ISTC/CSTIC 2009 (CISTC)

Editors:

D. Huang
Praxair
Indianapolis, Indiana, USA

C. Claeys
IMEC
Leuven, Belgium

P. Song
IBM
Yorktown Heights, New York, USA

C. Yu
Anji Inc.
Shanghai, China

T. Jiang
Micron
Boise, Idaho, USA

M. Kulkarni
MEMC
St. Peters, Missouri, USA

Y. Shi
Gritek
Beijing, China

R. Huang
Beijing University
Beijing, China

H. Wu
SMIC
Beijing, China

K. Wu
ASML
Shanghai, China

W. Wang
Intel
Leuven, Belgium

Q. Lin
IBM
Yorktown Heights, New York, USA

A. Peng
SEMI
Beijing, China



Published by
The Electrochemical Society
65 South Main Street, Building D
Pennington, NJ 08534-2839, USA
tel 609 737 1902
fax 609 737 2743
www.electrochem.org

ecs transactions ™

Vol. 18 No. 1

IPR2024-00533
Samsung Electronics Co. Ltd. et al v. Chien-Min Sung
Samsung's Exhibit 1016
Ex. 1016, Page 1

Copyright 2009 by The Electrochemical Society.
All rights reserved.

This book has been registered with Copyright Clearance Center.
For further information, please contact the Copyright Clearance Center,
Salem, Massachusetts.

Published by:

The Electrochemical Society
65 South Main Street
Pennington, New Jersey 08534-2839, USA

Telephone 609.737.1902
Fax 609.737.2743
e-mail: ecs@electrochem.org
Web: www.electrochem.org

ISSN 1938-6737 (online)
ISSN 1938-5862 (print)

ISBN 978-1-56677-703-2 (CD-ROM)
ISBN 978-1-60768-053-6 (PDF)
ISBN 978-1-60768-265-3 (Softcover)

Printed in the United States of America.

Preface

On behalf of the conference committee and organizers, we would like to thank you for your participation to one of the largest semiconductor conferences taking place in China, ISTC / CSTIC 2009. The 2009 conference is known as ISTC/CSTIC as a transition to CISTC (China International Semiconductor Technology Conference). CISTC is a merger between ISTC (International Semiconductor Technology Conference) and CSTIC (China Semiconductor Technology International Conference). CISTC will be the name for all conferences starting 2010 and beyond. It is organized jointly by ECS and SEMI, co-organized by CHTEC, and sponsored by IEEE and CEMIA. The 2009 conference will be the eighth international conference since ISTC was first initiated in 2001.

The mission of this conference is to provide a technical and professional forum for world experts to discuss both common and critical technologies, address the growing needs associated with silicon technology, and exchange their discoveries and solutions for current issues of high interest. We encourage collaboration, open discussion, and critical reviews at this conference. Furthermore, we hope that this conference will also provide collaborative opportunities for those who are interested in the semiconductor industry in Asia; particularly in China.

The conference committee consists of more than sixty world class experts from high profile companies including, but not limited to: Intel, IBM, SMIC, UMC, Micron, Motorola, SEMATECH, Applied Materials, IMEC, TEL, ASML, Praxair, MEMC, Cabot Microelectronics, Rohm&Haas, and also local companies and universities such as Anji, Gritek, Peking University, Fudan University and Zhejiang University. Guest speakers came from institutions mentioned above as well as renowned universities such as MIT, UC Berkeley, Stanford, Yale, Peking University, Tsinghua University, and other top universities around the world. We are proud of this prestigious volunteer team and feel honored by the distinguished speakers. With their contributions we believe that we have prepared a world class international conference.

The conference offers more than 270 high quality papers in nine symposia to cover most of the aspects of semiconductor technology. Symposium IX, Silicon Technology, will not only cover semiconductor materials, but will also provide key technology for solar applications. The conference started with a plenary session before all the symposia, covering a critical review and high level summary of the technology. Dr. Goerg Bednorz (Nobel Laureate/IBM Fellow), Dr. Robert Chau (Intel Sr. Fellow), Prof. Gilbert Declerck (CEO of IMEC), and Mr. Ray Roberge (CTO of Praxair) attended as distinguished conference keynote speakers. This volume collected most of the papers presented in the conference.

In spite of the many challenges this year in semiconductor industry, due in part to the world financial crisis, we have achieved a record high paper submission since ISTC started on 2001. We've seen strong participation from the US, Europe, areas in Asia, and other countries, and also noticed fast-growing paper submission from the host country of China. The following tables provide you with a brief summary of the paper distribution in each symposium and their geographical origin.

Table 1 – Conference Symposia and Papers

Symposium	Oral Presentation	Published in ECS	Poster
Plenary Session	4		
I - Design and Device Engineering	31	25	0
II - Metrology, Reliability and Testing	23	17	5
III - New Materials and Process Integration	9	6	3
IV - Photolithography	24	15	0
V - CMP & Cleaning	28	22	3
VI - Thin Film, Etch and Plating	27	18	10
VII - Packaging Technology	31	23	7
VIII - Emerging Semiconductor Technologies	31	11	2
IX - Silicon Technology	36	22	0
Total	**244**	**159**	**30**

Table 2 – Conference Paper Origin

Paper Origin	%
China	39%
US	30%
Europe	11%
Japan	8%
Taiwan	7%
Korea	3%
Singapore	2%

This conference would not have been a success without active participation from the audience and readers like you. We appreciate your support and trust that this issue of ECS *Transactions* will be a worthwhile addition and reference to your collection.

David Huang (Praxair Inc), Conference Chairman
Mark Ding, President of SEMI China
Roque J. Calvo, Executive Director of ECS
ISTC/CSTIC Committee
March 19, 2009, Shanghai, China

Editors (by Organization Committee)

Lead Editor
(Conference Chairman)

Dr. David Huang
Praxair
Indianapolis, Indiana, USA

Co-Editors
(Conference Co-Chairs and Symposium Chairs)

Conference Co-chair
Dr. Cor Claeys
IMEC
Leuven, Belgium

Conference Co-chair
Mrs. Ying Shi
Gritek
Beijing, China

Conference General Secretary
Mrs. April Peng
SEMI
Beijing, China

Symposium I Chairwoman
Prof. Ru Huang
Beijing University
Beijing, China

Symposium II Chairman
Dr. Peilin Song
IBM
Yorktown Heights, New York, USA

Conference Co-chair
Symposium III Chairman
Dr. Hanming Wu
SMIC
Beijing, China

Symposium IV Chairman
Dr. Ken Wu
ASML
Shanghai, China

Symposium V Chairman
Dr. Chris Yu
Anji Microelectronics
Shanghai, China

Symposium VI Chairman
Dr. Weie Wang
Intel
Leuven, Belgium

Symposium VII Chairman
Dr. Tom Jiang
Micron
Boise, Idaho, USA

Conference Co-chair
Symposium VIII Chairman
Dr. Qinghuang Lin
IBM
Yorktown Heights, New York, USA

Symposium IX Chairman
Dr. Milind Kulkarni
MEMC
St. Peters, Missouri, USA

Acknowledgement to Conference Committees, SEMI, and ECS Team

Conference Committees

Symposium I
Dr. Shaoning Mei
HHNEC, China

Dr. Boyong He
Shanghai Jiaotong University, China

Dr. Huiling Shang
IBM, USA

Dr. Hong Wu
SST, China

Symposium II
Dr. Wen-li Wu
NIST, USA

Dr. Yuhua Cheng
Peking University, China

Dr. Srinivas Raghvendra
Synopsys, USA

Dr. Timmy Bao
Veeco, USA

Dr. Teng Gao
IMEC, China

Dr. Francis Jen
KLA-Tencor, China

Symposium III
Dr. Youfan Liu
Intel, USA

Dr. Sowmya Krisnan
Ultra Clean Technology, USA

Dr. Chimin Sheu
Air Products, China

Dr. Ran LIU
Fudan University, China

Symposium IV
Dr. Kafai Lai
IBM, USA

Dr. Curtis Liang
ASML, China

Dr. Shuji Ding
Rohm & Haas Microelectronics Technologies, China

Prof. Zewen Liu
Tsinghua University, China

Dr. Zhibiao Mao
Toppan, China

Mr. Oliver Li
Intel

Dr. Gary Zhang
Rohm & Haas

Symposium V
Dr. Charles Xing
SMIC, China

Dr. Jingxun Fang
HHNEC, China

Dr. Yuling Liu
Hebei University of Technology, China

Dr. Kuochun Wu
Cabot Microelectronics, Asia

Dr. Christine Ye
Rohm & Haas Electronic Materials, USA

Symposium VI
Dr. Zhen Guo
Intel

Dr. Jason Tian
Nikon Precision, China

Dr. Canyuan Wang
Tokyo Electron, China

Dr. Ying Zhang
IBM, USA

Dr. Jerry Xu
Applied Materials, China

Symposium VII
Dr. Qunce Fu
Henkel

Dr. Ricky S W Lee
Hong Kong University of Science and Technology, China

Daniel Lu
Henkel

Symposium VIII
Dr. Jia Chen
IBM, USA

Dr. Fuhua Yang
Institute of Semiconductor, Chinese Academy of Sciences, China

Dr. Jie Zhang
Motorola, USA

Dr. Hyungjun Kim
Pohang University of Science and Technology, South Korea

Dr. Jinn P. Chu
National Taiwan University of Science and Technology, China

Dr. Wang Yueh
Intel, USA

Symposium IX
Dr. Renhe Jia
Applied Materials, USA

Dr. Deren Yang
Zhejiang University, China

SEMI Team
Polly Teng
April Peng
Susan Pan
Jessica Cao
Ivan Zhu

ECS Team
John Lewis
Amir Zaman
Beth Anne Stuebe

ECS Transactions, **Volume 18, Issue 1**
ISTC/CSTIC 2009 (CISTC)

Table of Contents

Preface *iii*

Chapter 1
Design and Device Engineering

Electrical Defect Issues of Hetero-Epitaxy for Advanced Nanometric CMOS Technologies * 3
C. Claeys, M. B. Gonzalez, G. Eneman, A. Hikavyy, R. Loo and E. Simoen (IMEC)

CMOS Compatible RF MEMS Switches with Serpentine Springs 9
X. L. Evans, H. Gamble, P. Baine, S. Mitchell, J. Montgomery, M. Bain, P. Rainey and D. Bien (Queen's University, Belfast)

Carbon co-implantation Used to Improve NMOS Threshold Voltage Roll-off Characteristics Study in 65nm Node CMOS Technology 15
J. Liu, Z. Allan, J. Jianhua, S. Dai, Z. Shen, E. Liu, B. Wang, A. Hung, H. Ho and J. Ning
(Semiconductor Manufacturing International Corporation)

Investigation into the Output Characteristics and Improvement of Operation Margin of IMOS (Impact Ionization MOS) Devices 21
H. Liu, R. Huang and Z. Wang
(Institute of Microelectronics, Peking University)

Minimizing Device Variation Caused by RTA Temperature Variation Across the Shot Field 27
J. Ju, E. Liu, S. Dai, A. Zhou, Z. Shen, J. Liu, B. Wang, D. Deng, J. Tang, A. Hung, H. Ho and J. Ning
(Semiconductor Manufacturing International Corporation)

Electrical Characterization of Mos Memory Devices with Self-Assembled Tungsten Nano-Dots Dispersed in Silicon Nitride 33
Y. Pei, C. Yin, M. Nishijima, T. Kojima (Tohoku University), H. Noriha (Musashi Institute of Technology), T. Fukushima, T. Tanaka and M. Koyanagi (Tohoku University)

Ex. 1016, Page 9

Electrical Characterization of W/Hfo2 MOSFETs with La2O3 Incorporation 39
H. Fujisawa (Tokyo Institute of Technology), A. Srivastava (Birla Institute of Technology and Science-PILANI), K. Kakushima, P. Ahmet, K. Tsutsui, N. Sugii, T. Hattori (Tokyo Institute of Technology), C. K. Sarkar (Jadavpur University) and H. Iwai (Tokyo Institute of Technology)

The Impact of Shallow Trench Isolation on the Channel Hot Electron Programming Behaviors of SONOS Cells 43
Z. Li, F. Yang, J. Wang, P. Lin, J. Chang, Y. Wang, C. Yi and F. Yan (Semiconductor Manufacturing International Corporation)

Bistable Resistance Switching of Cu/Ti:ZrO2/Pt for Nonvolatile Memory Application 49
Q. Liu, S. Long, Q. Wang, M. Liu, S. Zhang (Institute of Microelectronics, Chinese Academy of Sciences) and J. Chen (College of Electronics and Technology, Anhui University)

Transport in Silicon Nanowire Transistors * 55
T. Hiramoto, J. Chen, Y. Jeong and T. Saraya (University of Tokyo)

pMOSFET Off-state Leakage and Junction Leakage Current in Ge-based Devices 61
E. Simoen (IMEC), F. De Stefano (Universita degli Studi della Calabria, Italy), G. Eneman, B. De Jaeger, C. Claeys (IMEC) and F. Crupi (Universita degli Studi della Calabria, Italy)

A Study of Stress Technology for 65nm Generic CMOS Technology 67
X. Liu (Chartered Semiconductor MFG), W. Dai (UMC Singapore), X. Yu, L. Lim, B. Xue and B. He (CSM)

Evaluation of Lateral Ni Diffusion in Si Nanowire Schottky Contact 71
H. Kamimura, H. Arai, S. Sato, K. Kakushima, P. Ahmet, K. Tsutsui, N. Sugii, T. Hattori and H. Iwai (Tokyo Institute of Technology)

Characteristic Variations Induced by Random Fluctuation of Source/Drain Lateral Distribution Dopant in Nano-Scale UTB SOI MOSFET 77
L. Du, H. Deng, G. Du, R. Han (Institute of Microelectronics, Peking University) and S. Zhang (Peking University)

A Novel Accumulation Mode GAAC FinFET Transistor: Device Analysis, 3D TCAD Simulation and Fabrication 83
D. Xiao (Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences), M. Chi (Semiconductor Manufacturing International Corporation), D. Yuan (SYNOPSYS INC), X. Wang, Y. Yu (Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences), H. Wu and J. Xie (Semiconductor Manufacturing International Corporation)

Bit Cost Scalable (BiCS) Flash Technology for Future Ultra High Density Storage Devices * 89
A. Nitayama and H. Aochi (Toshiba Corp.)

Silicon Nanocrystals Synthesized by Electron-beam Co-evaporation Method and Their Application for Nonvolatile Memory 93
C. Chen, R. Jia, M. Liu, W. Li, C. X. Zhu, H. F. Li, P. W. Zhang, C. Q. Xie, Q. Wang and T. C. Ye
(Institute of Microelectronics, Chinese Academy of Sciences)

Investigations on the Impact of the Parasitic Bottom Transistor in Gate-All-Around Silicon Nanowire SONOS Memory Cells Fabricated on Bulk Si Substrate 99
Y. Ai, R. Huang, Y. Wang, J. Zhuge, D. Wu, R. Wang, P. Tang, L. Zhang, Z. Hao and Y. Wang (Institute of Microelectronics, Peking University)

Halo Implantation Optimization for 65nm Low Power PMOS Device Inverse Narrow Width Effect Improvement 105
H. Ye, J. Ju, X. Shi, B. Liu, J. Ning and I. C. Chen
(Semiconductor Manufacturing International Corporation)

A Novel Super-Halo Doping Concept for Nanoscale CMOS Technology 111
C. Liao (Shanghai Research Institute of Microelectronics) and S. Zhang (Peking University)

Shallow Trench Isolation Stress Effect on NMOS Transistor Leakage, SRAM Standby Current and VCCMIN 117
J. Ju, E. Liu, Z. Shen, A. Zhou, J. Liu, D. Deng, C. Ren, S. Wei, H. Ho and J. Ning (Semiconductor Manufacturing International Corporation)

A High Performance 80V Smart LDMOS Power Device Based on Thin SOI Technology 123
G. N. Sheu, Y. Hsu and S. Yang (Asia University)

Simulation Details for the Electrical Field Distribution and Breakdown Voltage of 0.15μThin Film SOI Power Device 129
H. N. You, Y. Liu, S. Tsaur and G. N. Sheu (Asia University)

Ex. 1016, Page 11

Investigation of Power MOSFET with Strained SiGe Channel 135
S. Sun (University of Central Florida), W. Zhou, L. Yan, J. Xu (Tsinghua National Laboratory for Information Science and Technology), J. Yuan and Z. Shen (University of Central Florida)

Simulation of Retention Characteristics for Metal Nanocrystal Nonvolatile Memory with a Modified Direct Tunneling Model 141
Y. Li, M. Liu, S. Long, Q. Wang, Q. Liu, S. Zhang, Y. Wang, Q. Zuo (Institute of Microelectronics, Chinese Academy of Sciences) and S. Liu (School of Physical Science and Technology, Lanzhou University)

Chapter 2
Metrology, Reliability and Testing

Metrology Requirements for the 32 nm Technology Node and Beyond * 151
J. Allgair (AMD Assignee to ISMI), B. Bunday, A. Cordes, P. Lipscomb, M. Godwin, V. Vartanian, M. Bishop, D. Arazi and K. Kim (ISMI)

Advanced Metrology for Nanoelectronics at the National Institute of Standards and Technology ** 161
J. V. Martinez de Pinillos (NIST), Y. Obeng and S. Knight (Office of Microelectronics Programs)

Advanced 3D-AFM Metrology for Sidewall Spacers 169
T. Bao, V. Ukraintsev and D. Dawson (Veeco Instruments)

High Yield Patterning with Simplified DFM Rules in Sub-100nm Process Technologies ** 177
L. Chang, C. Choi and M. Chi (Semiconductor Manufacturing International Corporation)

Evolution of Wafer Inspection and Review Methodology 183
E. Chang and A. Park (KLA-Tencor Corporation)

A Systematic Methodology For Low Yield Excursion Analysis 187
X. Xu, C. Lin, T. Tey and M. Hsu (United Microelectronics Corporation)

Use FPGA Configuration Memory For Backend Defect Capture 193
G. Qian and J. Shi (Xilinx Asia Pacific)

Functionality, Yield and Reliability Analysis of SiGe Micro-mirrors using Automated Optical Measurement Techniques * 199
I. De Wolf, J. De Coster, O. Varela Pedreira, L. Haspeslagh and A. Witvrouw (IMEC)

Reliability of Copper Interconnects: Stress-Induced Voids ** 205
J. Gambino, T. C. Lee, F. Chen and T. D. Sullivan (IBM)

Si-O Bond-breakage Energetics Under Consideration of the Whole Crystal 213
S. E. Tyaginov, V. Sverdlov, W. Gös, P. Schwaha, R. Heinzl, F. Stimpfl and T. Grasser (Institute for Microelectronics, TU Wien)

Drain Bias Effect on the Interface Traps of pMOSFETs under Negative Bias Temperature Stress 219
J. Pan, J. Yang, Y. Qiao, X. Liu, R. Han, J. Kang (Institute of Microelectronics, Peking University), C. C. Liao, H. Wu and Y. Wu (Semiconductor Manufacturing International Corporation)

Challenges of Contacting Lead-Free Packages 225
G. Gschwendtberger (Multitest Elektronische Systeme GmbH)

Power-aware Test Generation for Reducing Yield Loss Risk in At-Speed Scan Testing 231
Y. Yamato, X. Wen, K. Miyase, H. Furukawa and S. Kajihara (Kyushu Institute of Technology)

Yield Learning Through Scan Diagnosis ** 237
Y. Huang and W. Cheng (Mentor Graphics Co.)

A Low Power Architecture based on Unidirectional Cut 243
Z. Chen and D. Xiang (Tsinghua University)

A Self-Repairable Microprocessor 249
Y. Xie, X. Fu, Z. He (Institute of Computing Technology, Chinese Academy of Sciences), Y. Zhao (Department of Computer Science and Engineering, University of California, Riverside), Y. Hu and X. Li (Institute of Computing Technology, Chinese Academy of Sciences)

Improved Cost of Test by Optimized Tester Utilization 255
G. Jeserer (Multitest Elektronische Systeme GmbH)

Chapter 3
New Materials and Process Integration

Semiconductor Cleaning Technology -Past and Future ** 263
J. Ruzyllo (Penn State University)

Electromigration Improvement for Advanced Technology Nodes * 269
H. Wu, W. Wu, R. Shaviv, M. Sriram, A. Pradhan, K. Park, J. O'Loughlin, K. Chattopadhyay, T. Mountsier and G. Dixit (Novellus Systems)

Use of MIR-FTIR and k-value Measurements to Assess Potential Solutions to Reduce Damage During Porous Low-k Integration 275
J. Beynet, D. De Roest (ASM Belgium N.V.), N. Rochat (CEA, LETI, MINATEC), K. Kellens, P. Verdonck (IMEC) and H. Sprey (ASM Belgium N.V.)

Performance Improvement of CuOx RRAM for Non-volatile Applications * 281
Y. Lin, P. Zhou (Fudan University), M. Chi (Semiconductor Manufacturing International Corporation), H. B. Lv (Fudan University), S. Song (Semiconductor Manufacturing International Corporation), R. Huang, J. G. Wu, H. M. Wu (Memory Technology Development Center) and M. Chi (Semiconductor Manufacturing International Corporation)

Yield Improvement through Mechanism Study on WSi2 Abnormal Growth 287
J. Tang, Z. Fang and Z. Xu (Applied Materials China)

Optimization of Nickel Salicide Strip Process Using Taguchi Method 291
S. Yu, W. Li (UMC Singapore), C. Lin (United Microelectronics Corporation) and Y. Lu (UMC Singapore)

Chapter 4
Photolithography

Inverse Lithography Technology (ILT) Enabled Source Mask Optimization (SMO) ** 299
L. Pang, G. Xiao, V. Tolani, P. Hu, T. Cecil, T. Dam, K. Baik and B. Gleason (Luminescent Technologies)

Challenges and Opportunities for Optical CD Metrology in Future Lithography Process ** 315
J. Bao, J. Ferns and Y. Wen (Timbre Technologies, Inc.)

Line Edge Roughness Analysis for KrF Photoresists under sub 0.11 um Groundrules ** 321
L. Wang, L. Ji, P. Wu, H. Kan, W. He, Z. Zhu, F. Chen, F. Wei and J. Zhu (Shanghai Huahong NEC Electronics Company, Ltd.)

Lithography Process Transfer from ArF to High-NA KrF Tool ** 333
J. Hong (ASML) and A. Serebryakov (ABS)

Process Window Analysis for KrF Photoresists Under Sub-0.11um Groundrules 341
L. Wang, L. Ji, P. Wu, H. Kan, W. He, Z. Zhu, F. Chen, F. Wei and J. Zhu (Shanghai Huahong NEC Electronics Company, Ltd.)

Computation Lithography: Virtual Reality and Virtual Virtuality * 351
A. Wong (Magma Design Automation) and E. Lam (The University of Hong Kong)

OPC Accuracy and Development Time Improvement for Advanced Technology Nodes ** 357
P. Nikolsky, T. Nooitgedagt, P. Van Adrichem, J. Meessen, C. Kohler, B. Wisse (ASML Netherlands BV) and S. Hunsche (Brion, an ASML company)

Hot-Spots Aware Inverse Lithography Technology ** 367
Y. Yang, Z. Shi, S. Shen and C. Xie (Zhejiang University)

Trefoil Aberration Measurement of Lithographic Projection Optics Based on Linewidth Asymmetry of the Aerial Image 373
Q. Yuan, X. Wang and Z. Qiu (Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences)

High-Accuracy Measurement of Coma with the Optimized Alternating Phase-shifting Mask with a Specific Phase Width 381
Z. Qiu, X. Wang, Q. Yuan (Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences) and F. Wang (National Engineering Research Center for Lithographic Equipment)

EUV Source System Development for 22nm Generation and Beyond ** 391
B. Lin, D. Brandt and N. Farrar (Cymer Southeast Asia Ltd.)

Novel Embedded Barrier Layer Materials for ArF Non-topcoat Immersion Applications ** 397
D. Wang, C. J. Wu, C. Xu, G. Barclay, P. Trefonas and M. Xu (Rohm and Haas Electronic Materials)

Silicon-Containing Coupler Polymers Used For DUV Photoresist ** 405
R. Ran (CitiTech Advanced Materials Inc.)

A Cost Effective Spin on Sidewall Material Alternative to the CVD Sidewall Process ** 409
B. Ho, D. Maruyama, R. Sakamoto and Y. Sakaida (Nissan Chemical Industries)

Residue Testing of Developer-Soluble Bottom Anti-Reflective Coatings 419
C. Washburn, E. April and S. Brian (Brewer Science, Inc.)

Chapter 5
CMP & Cleaning

Investigation on Chemical Mechanical Polishing of GeSbTe for High Density Phase Change Memory 429
M. Zhong, Z. Song, B. Liu, S. Feng (Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences), Z. Xie, F. Zhang and Y. Xiang (Semiconductor Manufacturing International Corporation)

Study on CMP Mechanism and Technology of Sapphire Substrate for Photo-conducting Device 435
X. Niu, T. Wu, X. Zhao, B. Tan and Y. L. Liu
(Institute of Microelectronics, Hebei University of Technology)

Investigation of Defects in Post-CuCMP 441
S. Xiong, S. Huang, Y. Yang (Wuhan Xinxin Semiconductor Manufacturing Corporation), C. Tan, S. Wang, R. Zhao and J. Xu (Applied Materials China)

Study On The Electro-Chemical Cleaning Technique with BDD Film Anode to Preparation of Electronic Devices 447
B. Gao (Hebei University of Technology), Y. Zhu (Tianjin University of Technology), Y. Liu (Hebei University of Technology), X. Niu (Institute of Microelectronics, Hebei University of Technology), J. Zhang (Tianjin Polytechnic University), X. Liu and L. Li (Hebei University of Technology)

Cu/Low-k Thickness Measurement for Advanced Cu CMP Process Development and Control 453
Y. Li, N. Heylen, T. Delande, K. Kellens, P. Ong, L. Leunissen (IMEC), A. Tarnowka and A. Eliyahu (Nova Measuring Instruments Ltd.)

Cu/barrier CMP Protection Film Characterization Using Electrochemical Technique 459
S. Xu (Anji Microelectronics (Shanghai) Co., Ltd.)

Effect of Polisher Kinematics in Reducing Average and Variance of Shear Force and Increasing Removal Rate in Copper CMP 465
Y. Sampurno, A. Philipossian (University of Arizona), S. Theng (Araca Incorporated), T. Nemoto, X. Gu (Tohoku University), Y. Zhuang (Araca Incorporated), A. Teramoto and T. Ohmi (Tohoku University)

Single Dispersion to Polish both Bulk Cu and Residual Cu along with Barrier Layers 473
S. Janjam, D. Roy and S. V. Babu (Clarkson University)

Ultra-High Removal Rate Copper CMP Slurry Development for 3D Application ** 479
B. Hu, H. Kim, R. Wen and D. Mahulikar (Planar Solutions LLC)

Effect of H2O2 Concentration and Slurry Dilution on Removal Rate and Surface Quality for a Cu Slurry 485
J. Lin and T. Du (Fujimi Corporation)

Development of Copper CMP Slurry 491
J. Jing, C. Yang, X. Cai, D. Shiao and S. Wang (Anji Microelectronics (Shanghai) Co., Ltd.)

A New Kind of Technology for Post CMP Cleaning in IC Fabrication 497
L. Weiwei (Heibei University of Technology)

Controlled DI Water Gasification System for Advanced Semiconductor Cleaning Processes 503
A. Xia, K. Niermeyer, G. Conner and R. Mollica (Entegris Inc.)

A Novel Slurry for DRAM Polysilicon CMP 511
C. Wang (Anji Microelectronics (Shanghai) Co., Ltd.), Y. Cao (Semiconductor Manufacturing International Corporation), C. Yang (Anji Microelectronics (Shanghai) Co., Ltd.), H. Zhou (Semiconductor Manufacturing International Corporation), J. Jing (Anji Microelectronics (Shanghai) Co., Ltd.), Z. Xia, J. Tang, J. Lin, C. Chiu (Semiconductor Manufacturing International Corporation), S. Wang and C. Yu (Anji Microelectronics (Shanghai) Co., Ltd.)

A Matching Game for CMP Pads and Conditioners 517
F. F. Sun, J. Hawkins (Cabot Microelectronics Corporation), J. Tsai, G. Chiu (Cabot Microelectronics Corporation (Taiwan)) and A. Naman (Cabot Microelectronics Corporation)

Polycrystalline Diamond (PCD) Shaving Dresser: The Ultimate Diamond Disk (UDD) for CMP Pad Conditioning 523
J. C. Sung, C. Chou (Kinik Company), Y. Chen (National Defense University), C. Chou, Y. Pai, S. Hu (Kinik Company) and M. Sung (Advanced Diamond Solutions)

Development of CMP Conditioner for Sub-50 nm Technology Nodes 529
T. Hwang, R. Vedantham (Saint Gobain Abrasives, USA) and S. Zhang (Saint Gobain Abrasives, Shanghai)

The Design Features of Ultimate Diamond Disks (UDD): W CMP with In-Situ Dressing of Metal Free Diamond Disks 535
J. C. Sung, S. Hu, W. Huang, C. Chou, C. Chou and Y. Pai (Kinik Company)

Slurry Development Challenges and Solutions for Advanced Node Copper CMP ** 541
D. Mahulikar (Planar Solutions LLC)

Application of Surfactant in Si CMP Processing 547
B. Tan, X. Niu, P. Liang, T. Wu and Y. L. Liu (Institute of Microelectronics, Hebei University of Technology)

Poly-CMP Slurry for Flash Memories 553
J. Jing (Anji Microelectronics (Shanghai) Co., Ltd.), B. Chiu, T. Huang, A. Hsu (Applied Materials Taiwan), C. Yang and C. Yu (Anji Microelectronics (Shanghai) Co., Ltd.)

Synthesis of Spherical-like Ceria Particle with PVP as Assistant Agent and Its CMP Performance on Shallow Trench Isolation 559
Z. Zhang, W. Liu, Z. Song and X. Hu (Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences)

Chapter 6
Thin Film, Etch and Plating

Advanced Process Control of HfxSi1-xO2 Composition using Co-injection Atomic Layer Deposition 567
L. Bartholomew, A. Bavin, T. Lazerand, V. Rao (Aviza Technology), M. Weldon and L. Wang (Metrosol)

Thin Oxidation Study with Various Thermodynamic Parameters 575
P. Zhang, X. Zhang, T. Wen (North University of China), L. Xu, L. Shi (Southeast University), S. Webber, C. Willson (University of Texas at Austin) and K. Dean (SEMATECH International)

Solving Arcing Issues in CVD Processes 581
E. Yan, J. Xu, X. Li, J. Chen (Applied Materials China), Y. Chen, D. Zhao, L. Chan, L. Ji, D. Padhi and B. Zhang (Applied Materials)

Study on Bit Line TiCl4 TIN Process Optimization for Stack DRAM IDDS Standby Failure Improvement 587
L. Chen, Z. Li, X. Song, L. Xu, Y. Lin, H. Wei and P. Lin
(Semiconductor Manufacturing International Corporation)

Second Generation Copper Dielectric Barrier (BLOk II): k Reduction and Interface Control To Improve TDDB and EM ** 593
L. Xia, V. Zubkov, W. Ye, M. Shek, D. Cui and M. Naik
(Applied Materials, Inc.)

Studies of Wafer Backside Micro-arcing Prevention In Encore Ta Process 601
A. Chen, A. Bian, P. Lin, C. Xing and H. Zhou
(Semiconductor Manufacturing International Corporation)

Pre-clean Study for Both Damascene SiN and N-Blok to Resolve Queue-time-dependent Dielectric Barrier VBD Degradation 605
J. Xu, R. Gao, H. Wang, X. Li (Applied Materials China), S. Lee, D. Cui, M. Naik, L. Xia (Applied Materials, Inc.), Q. Xu, Z. Zhang and E. Bei
(Semiconductor Manufacturing International Corporation)

A Study on PVD Co Pre-clean Chamber RF Reflect Power Improvement 611
K. Gui, P. Hu, D. Ouyang, C. Peng, H. Yang and P. Lin
(Semiconductor Manufacturing International Corporation)

Robust HfxSi1-xO2 Gate Dielectric using Co-injection Atomic Layer Deposition 617
L. Bartholomew, A. Bavin, T. Lazerand and V. Rao (Aviza Technology)

Plasma Charging Damage Affecting Flash Memory Device Encountered at Pad Etch 625
S. Chong, K. A. Mohammad, F. Ng, L. J. Liu, D. Lee and S. F. Lee (X-Fab Sarawak Sdn. Bhd.)

Material Etch Rate Control in the Fluoride Containing Stripper for Post Etch and Ashing Residue Removal 635
B. Liu, L. Peng, J. Peng, J. Yu and S. Wang
(Anji Microelectronics (Shanghai) Co., Ltd.)

Influence of Polymeric Gas on Sidewall Profile and Defect Performance of Aluminum Metal Etch 641
X. Wang, H. Zhang, M. Shen, W. Sun and S. Chang
(Semiconductor Manufacturing International Corporation)

Etch Arts of Dual Damascus Structure and Their Impacts on WAT/VBD in 65nm Cu Interconnects 645
W. Sun, S. Chang, H. Zhang, X. Yin, L. Fu, B. Han, X. Wang and Y. Wu (Semiconductor Manufacturing International Corporation)

Understanding and Importance of Defects in Advanced Materials ** 651
L. Pantisano, L. Trojman, M. Aoulaiche (IMEC), R. O'Connor (Dublin City University), B. Kaczer and G. Groeseneken (IMEC)

Yield Enhancement with Optimized Offset Spacer Etch for 65nm Logic Low-leakage Process 659
B. Han, S. Chang, H. Zhang, L. Zhao, Y. Fu, W. Sun and B. Liu (Semiconductor Manufacturing International Corporation)

Study of Film Growth on Silicon in Advanced High Dose Implant Photoresist Strip 663
L. Diao (Mattson Technology) and S. Xu (Mattson Technology, Inc)

Via Resistance Variation Improvement Through Amine Versus Fluorine Based Chemical and O2 Dry Flushing for BiCMOS Technology 669
P. Tee, H. L. Teo, S. Liew, S. Chong, D. Lee and S. F. Lee (X-Fab Sarawak Sdn. Bhd.)

A Novel Approach for Electro Chemical Plating (ECP) Process Swirl Defect Reduction 677
R. Sun, J. Zhang, J. Jiang, Y. Bian, P. Lin, C. Xing (Semiconductor Manufacturing International Corporation), X. Li (Novellus System Inc.) and H. Zhou (Semiconductor Manufacturing International Corporation)

Chapter 7
Packaging Technology

Material Technologies for 3D Semiconductor Packaging * 683
N. Takano, T. Tanaka, S. Katogi and A. Nagai (Hitachi Chemical Co., Ltd.)

High-Speed Cu-TSV for 5:1 iTSV Applications ** 689
R. Beica and P. Siblerud (Semitool Inc.)

Advanced Technologies for In-Line and Post-Processed TSV Integration 695
E. O. Beyne, B. Swinnen, J. Van Olmen and D. S. Tezcan (IMEC)

3D DRAM in Server Application 701
B. Cai, V. Patel and E. D. Blackshear (IBM)

3D Integrated, Ultra-Thin Functional Microcontroller Device for Wireless, Flexible ECG Systems 707
W. Christiaens (Ghent University / IMEC), T. Torfs (IMEC), W. Huwel (ACB), C. Van Hoof (IMEC) and J. Vanfleteren (Ghent University / IMEC)

Challenges, Trends and Solutions for 3D Interconnects in Lithography and Wafer Level Bonding Techniques 713
M. Zoberbier (SUSS MicroTec Lithography), E. Hell (SUSS MicroTec Shanghai Co., Ltd.), K. Cook (SUSS MicroTec Inc.), M. Hennemayer and B. Neubert (SUSS MicroTec Lithography)

Throughput Enhanced Flip-Chip to Wafer Bonding: The Advanced Chip to Wafer Bonding 721
A. Sigl, S. Pargfrieder, T. Matthias, M. Wimplinger and P. Kettner (EV Group)

Material Solutions for Advanced Thin Packages 727
H. Tanaka (Sumitomo Bakelite Co. Ltd.)

Development of Epoxy-Silicone Hybrid Polymer and Application for Semiconductor Assembly 733
K. Tomiyoshi, T. Kanamaru and K. Fukui (Shin-Etsu Chemical Co., Ltd.)

UV Curable Adhesives for Cmos and Ccd Image Sensor Package 739
T. Kenmochi (Kyoritsu Chemical & Co., Ltd.)

Current Evolution of Wafer Thinning and Dicing 745
S. M. Sullivan (DISCO HI-TEC AMERICA, Inc.)

Electroless Nickel/Gold Under-Bump-Metallization Cost Comparisons 751
A. J. Strandjord, A. Scheffler (PacTech USA), E. Zakel, T. Oppert (PacTech GMBH) and T. Teutsch (PacTech USA)

Edge Protection of Temporarily Bonded Wafers during Backgrinding 757
D. Bai, X. Zhong, R. Puligadda (Brewer Science, Inc.), J. Burggraf, D. Burgstaller (EV Group), C. Lypka and J. Verzosa (DISCO HI-TEC AMERICA, Inc.)

Wafer Level SAW RF Filter Packaging With Through Wafer Via Interconnection 763
Y. Kweon, S. Park, T. Kim, J. Honh, S. Yang, J. Ha, T. Kim and S. Yi (Samsung Electromechanics)

High Volume, Low Cost Multi Die Packaging Using Conductive Polymer Interconnect 769
M. Robinson and J. Leal (Vertical Circuits, Inc.)

Next Generation Nickel-Based Bond Pads Enable Copper Wire Bonding 777
B. Chylak, J. Ling, H. Clauberg (Kulicke & Soffa Industries, Inc.) and T. Thieme (Atotech Deutschland GmbH)

Consideration of Gold Thickness in ENIG UBM Application 787
M. Nozu, T. Shibata, K. Yoshikawa and S. Hashimoto (C.Uyemura & Co., Ltd)

Dippable Paste: A Robust Assembly Material and Enabling Technology 793
H. G. Sy (Heraeus Materials Singapore Pte Ltd) and R. Lathrop (Heraeus Contact Division)

The Destructive Wire Bond Pull Test Methodology for the Ultra Low Loop Application on Thinner Die Package 803
O. Kwon (Kulicke & Soffa), B. Chylak (Kulicke & Soffa Industries, Inc.), O. Yauw and K. Lee (Kulicke & Soffa)

Grooved Ball Pad Structure (GBPS) to Improve FBGA Solder Joint Reliability (SJR) Performance 813
W. Zhou and T. Jiang (Micron Technology, Inc.)

Microdot Epoxy Stamping Process Qualification for Substrate-Based Optocoupler BGA Packaging 821
E. I. Livelo and R. V. Rojas (Fairchild Semiconductor)

Cu2 Process Development For Cost Effective SOT Package 831
Y. Soh and T. Lim (Infineon Technologies (Wuxi) Co., Ltd.)

Enabling Technology in Thin Wafer Dicing 837
J. Van Borkulo, R. Evertsen and R. Hendriks (Advanced Laser Separation International ALSI)

Chapter 8
Emerging Semiconductor Technologies

Growth and Electrical Characterization of Horizontally Aligned CNTs 845
C. Santini (IMEC / University of Palermo), D. J. Cott, A. Romo Negreira (IMEC), S. Riva Sanseverino (University of Palermo), S. De Gendt (IMEC / KU Leuven Chemistry Department) and P. M. Vereecken (IMEC)

Fine Pitch Transfer and Join (T&J) From MCM to 3DI ** 851
R. R. Yu (IBM)

Exploratory NEMS-CMOS Hybrid Devices for Post CMOS Era ** 857
B. Lee (Gwangju Institute of Science and Technology)

Evaluation of InAs QWFET for Low Power Logic Applications ** 863
E. Chang, C. Kuo (National Chiao Tung University), H. Hsu (Yuan Ze University), Y. Miyamoto (Tokyo Institute of Technology), C. Chang and C. Wu (National Chiao Tung University)

Translucent Polymer Solar Cells ** 869
G. Ng, C. Vijila, W. Goh and F. Zhu
(Institute of Materials Research and Engineering)

PV Efficiency Improved Large-area Dye-sensitized Solar Cells 875
H. Yin, T. Zhang, X. Wang, L. Liu, Y. Wang, X. Liu and J. Kang
(Institute of Microelectronics, Peking University)

Influence of Sulfurization Temperature on Structure and Semiconducting Properties of SnxSy Thin Films 881
S. Cheng, S. Peng and Y. Yang (Fuzhou University)

Thin Film InP Epitaxy on Si (001) Using Selective Aspect Ratio Trapping 887
J. Li, J. Bai, J. M. Hydrick, J. G. Fiorenza, C. Major, M. Carroll (Amberwave), Z. Shellenbarger (Sarnoff Corporation) and A. Lochtefeld (Amberwave)

Sub-micrometer Organic Field Effect Transistors 895
L. Shang, M. Liu, Z. Ji, G. Liu, X. Liu, J. Liu and H. Wang
(Institute of Microelectronics, Chinese Academy of Sciences)

Resistive Switching Characteristics of MnOx-based ReRAM 901
S. Zhang, S. Long, Q. Liu, Q. Wang and M. Liu
(Institute of Microelectronics, Chinese Academy of Sciences)

FEA Based Design Optimization of Exciting Sensitivity for Micromachined Beam Ultrasonic Transducer 907
Y. Han (Nokia Research Center), C. Cai, Y. Liu, A. Salo, L. Xu (Beijing Jiaotong Univeresity) and J. Xie (Nokia Research Center)

Chapter 9
Silicon Technology for Semiconductor, Electronic, and Solar Applications

Polysilicon Production: Reaction Engineering and Scaleup Issues * 915
P. A. Ramachandran (Washington University)

Heat and Impurity Transfer Mechanisms of Czochralski and Directional Solidification Processes * 925
K. Kakimoto, X. Chen, L. J. Liu, H. Miyazawa, H. Matsuo, S. Nakano, S. Hisamatsu and Y. Kangawa (Kyushu University)

Global Simulation of CZ and FZ Bulk Crystal Growth: From Quasi-Dynamic and Dynamic Modelling to Process Control and Crystal Quality Optimization * 935
F. Dupret (Universite Catholique de Louvain), R. Rolinsky, L. Wu, F. Loix, A. De Potter, N. Van den Bogaert and V. Regnier (FEMAGSoft S.A. Company)

Agglomeration of Vacancies into Voids and Oxide Particles in Silicon Crystals * 945
V. V. Voronkov (MEMC Electronic Materials)

Atomistic-to-Continuum Modeling of Defect-Related Phenomena in Silicon Crystals * 959
T. Sinno and S. S. Kapur (University of Pennsylvania)

Challenges in Wafering Processes for Semiconductor and Solar Wafers * 969
H. Erk (MEMC Electronic Materials, Inc)

Prediction of Melt-Crystal Interface Shape and Melt Convection in a Large-scale CZ-Si Growth System Using RANS and LES Methods in Global Simulation. 983
X. Liu, L. Liu, Z. Li, Y. Wang (School of Energy and Power Engineering, Xi'an Jiaotong University) and K. Kakimoto (Kyushu University)

Calibration of the Enthalpy-Porosity Based Method for CZ Silicon Growth 989
B. Dai, B. Devulapalli (MEMC Electronic Materials Inc.), P. Gunjalb (Tridiagonal Solutions) and M. Kulkarni (MEMC Electronic Materials Inc.)

Advances in the Understanding of Oxide Precipitate Nucleation in Silicon ** 995
G. Kissinger, D. Kot, V. Akhmetov (IHP), A. Sattler, T. Mueller and W. Von Ammon (Siltronic AG)

Oxygen Precipitation in Heavily Phosphorus-doped Czochralski Silicon 1001
Y. Zeng, D. Yang, X. Ma, X. Zhang, L. Lin and D. Que (State Key Laboratory of Silicon Materials)

Interactions of Cu and Ni Impurities with Vacancy-related Point Defects in Czochralski-grown Si Crystals 1013
V. Markevich, A. R. Peaker (University of Manchester), I. F. Medvedeva, V. E. Gusakov, L. I. Murin (Scientific-Practical Materials Research Centre of NAS of Belarus) and B. G. Svensson (University of Oslo)

Effect of Oxygen Precipitates in Annealed Electron Irradiated Czochralski Silicon 1019
G. Chen, X. Ma, Y. Bai, L. Cai and Y. Li (Hebei University of Technology)

Modeling Multi-Crystalline Silicon Growth in Directional Solidification Systems 1023
B. Devulapalli and M. Kulkarni (MEMC Electronic Materials Inc.)

Simulation Analysis of Point Defects in Silicon Ingot during Unidirectional Solidification Process for Solar Cells 1031
X. Chen, S. Nakano (Kyushu University), L. Liu (School of Energy and Power Engineering, Xi'an Jiaotong University) and K. Kakimoto (Kyushu University)

Distributions of Light Elements and Their Precipitations Grown by Unidirectional Solidification Method in Multicrystalline Silicon for Solar Cells 1037
H. Matsuo, S. Hisamatsu, Y. Kangawa and K. Kakimoto (Kyushu University)

Numerical Simulation of Thermal Distribution during the Directional Solidification of Polycrystalline Silicon Ingot 1043
D. Wu, W. Sun, Q. Hao, J. Zhao and C. Liu (Hebei University of Technology)

Performance Limitations of mc-Si Solar Cells Caused by Defect Clusters ** 1049
B. Sopori, P. Rupnowski, V. Mehta, V. Budhraja, S. Johnston, N. Call, H. Mountinho, M. Al-Jassim (National Renewable Energy Laboratory), A. Shaikh (Ferro Electronic Materials), M. Seacrist (MEMC Electronic Materials) and D. Carlson (BP Solar)

Cluster Ion Implantation System: Claris for Beyond 45nm Device Fabrication (Ii) 1059
M. Tanjyo, N. Hamamoto, T. Nagayama, S. Umisedo, Y. Koga, N. Maehara, H. Une, T. Matsumoto, N. Nagai (Nissin Ion Equipment Co.,Ltd.) and J. O. Borland (JOB Technology Inc.)

Silicon Etching Technology Using Inductively Coupled Plasma Suitable for Fabrication of Silicon Nanowires and Ring Resonators 1065
J. Zhou, H. Shen, H. Zhang and X. Liu
(Institute of Microelectronics of Chinese Academy of Sciences)

The Synthesization of Si Nanocrystals Embedded in HfO2 Matrix by Electron Beam co-Evaporation Technique 1071
W. Li, R. Jia, C. Chen, M. Liu, H. F. Li, C. X. Zhu and S. Long
(Institute of Microelectronics, Chinese Academy of Sciences)

New Structure Improves Copper Phthalocyanine (CuPc) Order and Decrease the Device Resistance 1075
G. Liu, M. Liu, L. Shang, Z. Ji, X. Liu, H. Wang and J. Liu
(Institute of Microelectronics, Chinese Academy of Sciences)

Investigating the Effects of Different Gate Biasing Voltages and Temperatures on RadFETs 1081
H. Tang, Y. Liu, Y. Wang, J. Wang and Y. Jin
(Institute of Microelectronics, Peking University)

Author Index 1087

** Keynote paper*
*** Invited paper*



Facts about ECS

The Electrochemical Society (ECS) is an international, nonprofit, scientific, educational organization founded for the advancement of the theory and practice of electrochemistry, electrothermics, electronics, and allied subjects. The Society was founded in Philadelphia in 1902 and incorporated in 1930. There are currently over 7,000 scientists and engineers from more than 70 countries who hold individual membership; the Society is also supported by more than 100 corporations through Corporate Memberships.

The technical activities of the Society are carried on by Divisions. Sections of the Society have been organized in a number of cities and regions. Major international meetings of the Society are held in the spring and fall of each year. At these meetings, the Divisions and Groups hold general sessions and sponsor symposia on specialized subjects.

The Society has an active publications program that includes the following.

Journal of The Electrochemical Society — JES is the peer-reviewed leader in the field of electrochemical and solid-state science and technology. Articles are posted online as soon as they become available for publication. This archival journal is also available in a paper edition, published monthly following electronic publication.

Electrochemical and Solid-State Letters — ESL is the first and only rapid-publication electronic journal covering the same technical areas as JES. Articles are posted online as soon as they become available for publication. This peer-reviewed, archival journal is also available in a paper edition, published monthly following electronic publication. It is a joint publication of ECS and the IEEE Electron Devices Society.

Interface — *Interface* is ECS's quarterly news magazine. It provides a forum for the lively exchange of ideas and news among members of ECS and the international scientific community at large. Published online (with free access to all) and in paper, issues highlight special features on the state of electrochemical and solid-state science and technology. The paper edition is automatically sent to all ECS members.

Meeting Abstracts (formerly Extended Abstracts) — Abstracts of the technical papers presented at the spring and fall meetings of the Society are published on CD-ROM.

ECS Transactions — This online database provides access to full-text articles presented at ECS and ECS-sponsored meetings. Content is available through individual articles, or as collections of articles representing entire symposia.

Monograph Volumes — The Society sponsors the publication of hardbound monograph volumes, which provide authoritative accounts of specific topics in electrochemistry, solid-state science, and related disciplines.

For more information on these and other Society activities, visit the ECS website:

www.electrochem.org

CHAPTER 1

DESIGN AND DEVICE ENGINEERING

ECS Transactions, 18 (1) 3-8 (2009)
10.1149/1.3096418 © The Electrochemical Society

Electrical Defect Issues of Hetero-Epitaxy for Advanced Nanometric CMOS Technologies

C. Claeys[a,b], M. Bargallo Gonzalez[a,b], G. Eneman[a,b,c], A. Hikavyy[a], R. Loo[a], and E. Simoen[a]

[a]IMEC, Kapeldreef 75, B-3001 Leuven, Belgium
[b]EE Depart., KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
[c]Post-doc Researcher Fonds Wetenschappelijk Onderzoek Vlaanderen

Epitaxial techniques have been around for several decades already, but are now becoming an essential process module for several state-of-the-art process technologies, whereby the stringent defect control requirements have imposed new challenges. Some state-of-the-art hetero-epitaxial applications are briefly reviewed from a defect engineering point of view.

Introduction

The early use of Si epitaxy was aiming at enhancing the performance of discrete bipolar devices by achieving high-resistive layers to control the breakdown voltage [1]. Later on epitaxial layers were extensively used in bipolar integrated circuits for e.g. controlling collector regions, buried layers and isolation regions. In the 80-ties, the homo-epitaxy of a Si film on a silicon substrate was considered as the solution to solve latch-up problems in CMOS technologies [2]. The introduction of trench isolation reduced the latch-up concern so that the main application field of epitaxial material then became related to the improvement of the Gate-Oxide-Integrity (GOI) and the DRAM refresh properties (see e.g. [3] and references therein), although cost considerations remained important. Since the 90-ties there is a strong interest in SiGe epitaxy for BiCMOS and rf applications [4], as high-mobility channel material, and for optical applications. The drive for boosting up the CMOS device performance has strongly triggered the use of hetero-epitaxy for either the fabrication of a high-mobility channel material or for strain engineering based on the selective growth of compressive (SiGe) or tensile (SiC) strained source/drain regions [5], while other recent applications are using the growth of strained Si (or Ge) on a relaxed $Si_{1-x}Ge_x$ buffer layer, introducing biaxial strain from the bottom substrate. An overview of some epitaxial techniques used for strain engineering can be found in [6] and are schematically shown in Fig. 1.

The quest for increased drive current and carrier mobility has lead to a revival of the use of Ge either as bulk wafers or for thin film applications such as Ge-on-Silicon or Ge-on-insulator [7]. In both cases hetero-epitaxy plays an important role. Although very promising p-channel Ge devices have been reported, controlling n-channel Ge devices remains difficult. Therefore, there is strong interest in using III-V materials for the latter. Extensive research has been focusing on the monolithic integration of III-V semiconductors with Si. However, excessive defect nucleation associated with the large lattice mismatch with Si may occur and degrade the device performance [8]. In recent years, promising results have been achieved by Fitzgerald and co-workers (see e.g. [9,10]) using a $Si_{1-x}Ge_x$ buffer layer with a gradually changing composition. As will be discussed further in more detail, a low threading dislocation density (TDD) is essential for controlling the minority carrier lifetime and the leakage current of the devices. Thanks

Ex. 1016, Page 31

to the progress in epitaxial techniques, the ultimate monolithic integration of both Ge and III-V devices on a silicon substrate seems to become within reach and has triggered world-wide research activities. In Europe the research in this field is coordinated by the DUALLOGIC project [11].

One of the key processing issues is the control of the defect generation associated with the epitaxial processes. Too high stresses, related to the lattice mismatch, will initiate relaxation processes resulting in the nucleation of misfit and threading dislocations. These dislocations cannot only affect physical processes such as e.g. dopant diffusion but will also influence several electrical performance parameters like leakage current, carrier lifetime, low frequency noise, etc.

The present paper would like to review important defect aspects for some state-of-the-art hetero-epitaxial applications.

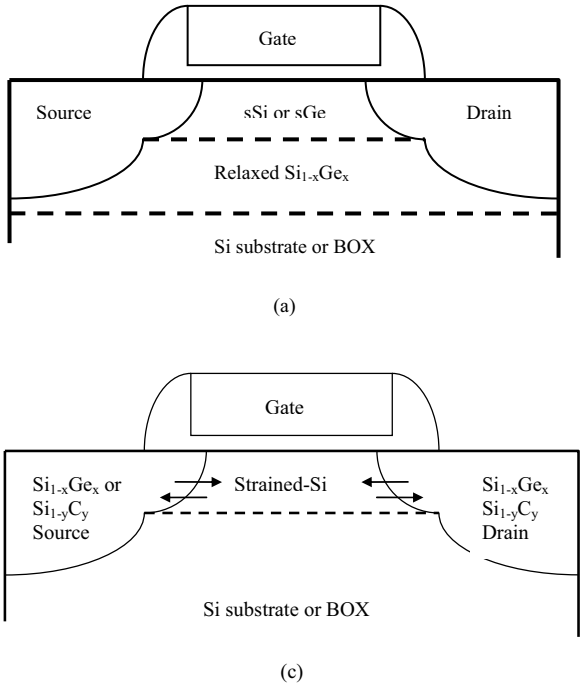


Figure 1. Schematic representation of different strain-engineering approaches: (a) global strain, using a tensile strained-Si (sSi) (x<0.5) or compressively strained-Ge (sGe) layer (x>0.5) on a silicon or Silicon-on-Insulator substrate. The sSi or sGe layer may be deposited directly on the buried oxide (BOX); (b) local tensile or compressive stress induced by a SiN cap layer in the Si, Ge or $Si_{1-x}Ge_x$ channel; (c) local tensile or compressive strain induced by embedded $Si_{1-x}Ge_x$ or $Si_{1-y}C_y$ source/drain regions.

Embedded $Si_{1-x}Ge_x$ and $Si_{1-y}C_y$ Source/Drain Regions

Local hetero-epitaxial growth allows the formation of embedded source/drain regions, as illustrated in Fig. 1c, whereby either tensile or compressive stress is built-in depending on whether SiGe or SiC is used. There are a large variety of technological parameters (Si etch depth, degree of recess, epitaxial thickness, layer composition, pre-epitaxial cleaning and bake, growth conditions, anneal technique, junction geometry, use of pre- or post-epi implants, etc) that will have an impact on the final stress level and the associated defect generation [12-17]. Optimization requires a subtle interplay between the different parameters and is essential for controlling important effects such as band gap engineering, carrier mobility enhancement, defect control, leakage current reduction, lifetime improvement, self-heating reduction, etc. The goal is to achieve structures with the

Ex. 1016, Page 32

highest possible stress level for carrier mobility enhancement, without the occurrence of stress relaxation and/or extended defects formation. In the initial phase of stress relaxation, the remaining stress in the depletion region can be sufficiently high for the band gap narrowing induced leakage current to dominate, before the relaxation-induced defects reduce the carrier lifetime enhancing the area leakage current [17]. A schematic illustration of the different area leakage current density regimes versus the stress level (or critical epitaxial thickness) is shown in Fig. 2.

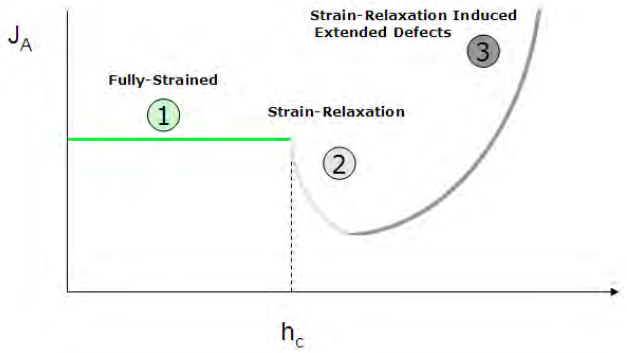


Figure 2. Schematic illustration of the area leakage current regimes as a function of the thickness of the embedded epitaxial layer.

A theoretical analysis of the stress-induced leakage current has to take into account that a compressive (tensile) biaxial in-plane stress in the $Si_{1-x}Ge_x$ ($Si_{1-y}C_y$) epitaxial layer will translate into a biaxial tensile (compressive) stress in the upper part of the silicon substrate, adjacent to the SiGe/Si (SiC/Si) heterojunction. Therefore, the total generation current is dominated by the biaxial stress present in the top of the depletion region.

For these embedded source/drain structures the stress in the material has been extensively studied both experimentally (Raman measurements) and TSUPREM4 2D simulations [18]. An illustration is given in Fig. 3a, showing for 20% Ge the simulated stress relaxation as a function of the active area. The strong area dependence for window sizes below 5 µm is confirmed by HR-XRD investigations as shown in Fig. 3b [19].

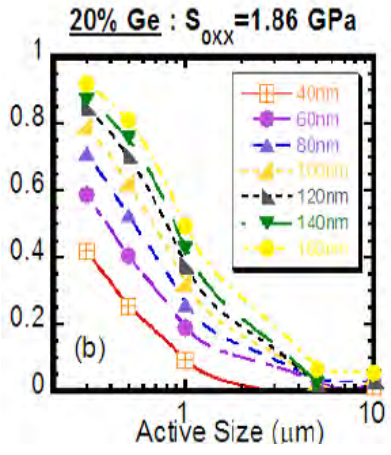


Figure 3a. TSUPREM4 2D simulation of the elastic relaxation as function of active size for the σ_{xx} stress component in the [110] direction for 20% Ge. The stress values were taken in the middle of the layer.

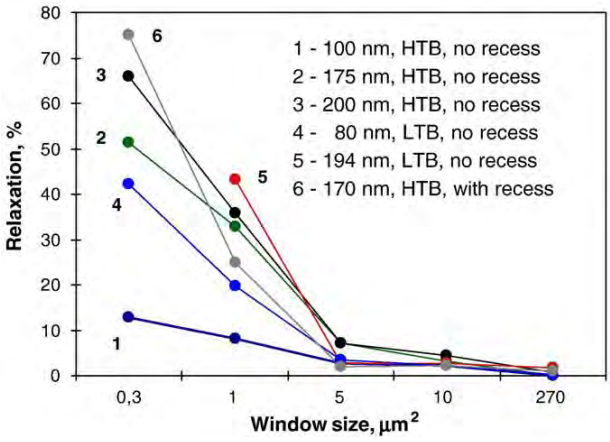


Figure 3b. Relaxation as function of window size determined by HR-XRD for different SiGe layers (80-200 nm, high (HTB) or low temperature pre-bake (LTB), with or without Si recess).

Epitaxial Si as Passivation Layer for Ge Devices

The search for improved carrier mobility in sub 32 nm CMOS technologies has worldwide triggered the renewed interest in using Ge devices [7]. Ge films on insulator (GeOI) can be fabricated by either wafer bonding (used for thicker layers), the smart-cut process (enables thin layers) or the Ge condensation method. In the latter case a SiGe layer is epitaxially grown on a SOI substrate and subsequently oxidized so that the SiGe layer is converted into a Ge-rich layer with a SiO_2 layer on top which can be etched away. It is also possible to grow epitaxial Ge directly on the Si substrate, but the main problem is to obtain a low threading dislocation density. Extended defect formation in Ge has recently been reviewed in detail [20]. Nowadays, defect engineering aspects related to substrate fabrication, shallow junction formation and germanidation are well understood [21]. Important remaining Ge-processing challenges are interface passivation, gate stack formation, contact technology, and the fabrication of high-performing n-channel devices.

Record performances of I_{on} = 478 μA/μm and I_{off} = 37 nA/μm at V_{DD}= -1 V have been reported for 65 nm Ge pMOSFETs [22]. A key element is the gate stack consisting of a few monolayers of Si, a 0.5 nm SiO_2 and a 4 nm ALD HfO_2. The thickness of the Si passivation layer, epitaxially grown on the Ge, has to be critically controlled as a too thin layer causes problems with the exposure of the Si-Ge interface during the oxidation and a too thick layer gives stress enhancing the density of interface states [23]. Figure 4 illustrates some cross-sectional TEM images of high-k gate stacks in Ge with different thickness of the Si passivation layer. It can clearly be seen that for 12 MLs the mismatch stress is relaxed by the generation of misfit dislocations.

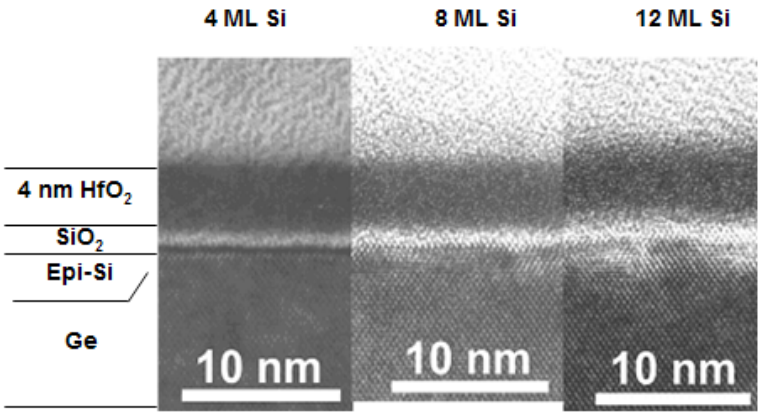


Figure 4. Cross-sectional TEM images of a high-k gate stack on a Ge surface passivated by different thicknesses of Si. At 12 MLs of silicon, the layer relaxes, giving rise to misfit dislocations at the interface.

For an optimal epitaxial growth the surface preparation is critical, i.e., the use of an HF dip followed by an *in-situ* H_2 anneal [23]. Another important parameter is the deposition temperature as a too high temperature leads to Ge surface segregation, thereby increasing the density of interface traps [243]. Although Si passivation works well for Ge p-MOSFETs, alternative surface passivation techniques are under investigation for n-channel devices.

Monolithic Integration on Lattice Engineered Si Substrates

As mentioned in the introduction, there is a great interest in the monolithic integration of III-V materials on Si, with as final goal the co-integration of III-V and Ge components with Si CMOS to fabricate systems-on-chip applications. Therefore, defect engineering is a great challenge for hetero-epitaxial processes. A lot of the earlier work in the field has been driven by III-V multijunction solar cell technology on Si applications such as e.g. $In_{0.49}Ga_{0.51}P$/GaAs [254] or GaInP/GaAs [26] dual junction cells on compositional graded $Si_{1-x}Ge_x$/Si. For these applications TDDs below 10^6 cm^{-2} are achievable, resulting in GaAs lifetimes below 10 ns. Graded SiGe buffer layers can also be used for high gain AlGaAs/GaAs bipolar devices (emitter area $1.6x10^3$ μm^2 and base concentration of $1x10^{19}$ cm^{-3}) with peak dc current gains above 100 [27]. To characterize the quality of the epitaxial layers, p-n junctions are highly sensitive device structures [28].

Conclusions

Hetero-epitaxy has become an essential process module for state-of-the-art CMOS technologies and offers a great potential for future applications based on the monolithic integration of III-V on Si. Defect engineering, however, will be crucial for achieving the required device performance.

Acknowledgments

We would like to acknowledge the IMEC epi and the Ge/III-V Team for helpful discussions and the use of co-authored results.

References

1. H.C. Theuerer, J.J. Kleimack, H.H. Loar and H. Christensen, *Proc. IRE*, **48**, 1642 (1960)
2. R.R. Troutman, *Latch-up in CMOS Technology – The Problem and Its Cure*, Kluwer Acad. Press. (1986)
3. H.R. Huff, in *ULSI Science and Technology 1997,* Eds M.Z. Massoud, H. Iwai, C. Claeys and R. Fair, Electrochem Soc. Proc. Vol. **PV 97-3**, 53 (1997)
4. J.D. Cressler, *Silicon Heterostructure handbook: Materials, Fabrication, Devices, circuits, and Applications of SiGe and Si Strained-Layer Epitaxy*, Ed. J.D. Cressler, CRC Press (2006)
5. G. Eneman, E. Simoen, P. Verheyen and K. De Meyer, *IEEE Trans. Electron Dev.*, 55, 2703 (2008)
6. C. Claeys, E. Simoen, S. Put, G. Giusi and F. Crupi, *Solid-State Electron.*, **52**, 1115 (2008)
7. C. Claeys and E. Simoen, *Germanium-Based Technologies: From Materials to Devices*, Elsevier (2007)
8. M. Yamaguchi and C. Amano, *J. Appl. Phys.*, **58**, 3601 (1985)
9. E.A. Fitzgerald, Y.-H. Xie, D. Monroe, P.J. Silverman, J.M. Kuo, A.R. Kortan, F.A. Thiel and B.E. Weir, *J. Vac. Sci. Technol. B*, **10**, 1807 (1992)
10. R.M. Sieg, J.A. Carlin, J.J. Boeckl, S.A. Ringel, M.T. Currie, S.M. Ting, T.A. Langdo, G. Taraschi, E.A. Fitzgerald and B.M. Keynes, *Appl. Phys. Lett.*, **73**, 3111 (1998)
11. http://www.ims.demokritos.gr/DUALLOGIC

12. C. Claeys, M. Bargallo Gonzalez, G. Eneman, P. Verheyen, H. Bender, R. Schreutelkamp, L. Washington, F. Nouri and E. Simoen, *J. Electrochem. Soc*., **154**, H814 (2007)
13. M.B. Gonzalez, M.K. Chowdhury, N. Bhouri, P. Verheyen, F. Leys, O. Richard, R. Loo, C. Claeys, E. Simoen, V. Machkaoutsan, P. Tomasini, S.G. Thomas, J.P. Lu, J.W. Weijtmans and R. Wise, *Electrochem. Soc. Trans*., **6(1)**, 389 (2007)
14. M. Bargallo Gonzalez, N. Thomas, E. Simoen, P. Verheyen, A. Hikavyy,F.E. Leys, Y. Okuno, B. Vissouvanadin, B. Van Daele, L. Geenen, R. Loo, C. Claeys, V. Machkaoutsan, P. Tomasini, S.G. Thomas, J.P. Lu, J.W. Weijtmans and R. Wise, *Electrochem. Soc. Trans*., **11(3)**, 47 (2007)
15. M.K. Chowdhury, B. Vissouvanadin, M.B. Gonzalez, N. Bhouri, P. Verheyen, A. Hikavyy, O. Richard, J. Geypen, H. Bender, R. Loo, C. Claeys, E. Simoen, V. Machkaoutsan, P. Tomasini, S.G. Thomas, J.P. Lu, J.W. Weijtmans and R. Wise, *Solid State Phenom.,* **131-133**, 95 (2008)
16. E. Simoen, M. B. Gonzalez, B. Vissouvanadin, M.K. Chowdhury, P. Verheyen, A. Hikavyy, H. Bender, R. Loo, C. Claeys, V. Machkaoutsan, P. Tomasini, S. Thomas, J.P. Lu, J.W. Weijtmans and R. Wise, *IEEE Trans. Electron Dev*., **55**, 925 (2008)
17. M.B. Gonzalez, E. Simoen, B. Vissouvanadin, P. Verheyen, R. Loo and C. Claeys, *IEEE Trans. Electron Dev.* (submitted)
18. M.B. Gonzalez, E. Simoen, B. Vissouvanadin, G. Eneman, P. Verheyen, R. Loo, C. Claeys, V. Machkaoutsan, P. Tomasini and S. Thomas, *Proc. E-MRS* (accepted)
19. A. Hikavyy, N. Bhouri, R. Loo, P. Verheyen, F. Clemente, J. Hopkins, R. Trussell and M. Caymax, *Thin Solid Films*, **517**,113 (2008)
20. C. Claeys and E. Simoen, *Fundamental and Technological Aspects of Extended Defects in Germanium*, Springer-Verlag (2009)
21. C. Claeys, E. Simoen, K. Opsomer, D.P. Brunco and M. Meuris, *Mat Sci Semicond Process.*,**154-155**, 49 (2008).
22. J. Mitard, B. De Jaeger, F.E. Leys, G. Hellings, K. Martens, G. Eneman, D.P. Brunco, R. Loo, J.C. Lin, D. Shamiryan, T. Vandeweyer, G. Winderickx, E. Vrancken, C.H. Yu, K. De Meyer, M. Caymax, L. Pantisano, M. Meuris and M.M. Heyns, *IEDM Techn. Digest*, 873 (2008)
23. B. De Jaeger, R. Bonzom, F. Leys, O. Richard, J. Van Steenbergen, G. Winderickx, E. Van Moorhem, G. Raskin, F. Letertre, T. Billon, M. Meuris and M.M. Heyns, *Microelectron. Engrg*., **80**, 26 (2005)
24. F.E. Leys, R. Bonzom, B. Kaczer, T. Janssens, W. Vandervorst, B. De Jaeger, J. Van Steenbergen, K. Martens, D. Hellin, J. Rip, G. Dilliway, A. Delabie, P. Zimmerman, M. Houssa, A. Theuwis, R. Loo, M. Meuris, M. Caymax, M.M. Heyns, *Mat. Sci*. *Semicond. Process*., **9**, 679 (2006)
25. S.A. Ringel, C.L. Andre, E.A. Fitzgerald, A.J. Pitera and D.M. Witt, in *Proc. 16th Eur. Photov. Solar Energy Conf*., Glasgow, 567 (2000)
26. M.R. Lueck, C.L. Andre, A.J. Pitera, M.L. Lee, E.A. Fitzgerald and S.A. Singel, *IEEE Electron Dev. Lett*., **27**, 142 (2006)
27. K.L. lew, S.F. Yoon, W.K. Loke, H. Tanoto, C.L. Dohrman, D.M. Issacson and E.A. Fitzgerald, *J. Vac. Sci. Technol*. B, **25**, 902 (2007)
28. K.P. Chen, S.F. Yoon, T.K. Ng, H. Tanoto, K.L. Lew, C.L. Dohrman and E.A. Fitzgerald, *J. Appl. Phys*., **104**, 073710 (2008)

ECS Transactions, 18 (1) 9-14 (2009)
10.1149/1.3096419 © The Electrochemical Society

CMOS Compatible RF MEMS Switches with Serpentine Springs

X.L. Evans, H.S. Gamble, P.T. Baine, S.J.N. Mitchell, J. Montgomery, M.F. Bain, P.V. Rainey, and D.C.S Bien

School of Electronics, Electrical Engineering and Computer Science, Queen's University, Belfast, Northern Ireland BT9 5DR, U.K.

This paper describes a serpentine flexure spring design and fabrication process development for radio frequency microelectromechanical (RF MEMS) capacitive switches with coplanar waveguide (CPW) lines. Sputtered tungsten is employed as the CPW line conductor instead of Au, a non-Si compatible material. The bridge membrane is fabricated from Al. The materials and fabrication process can be integrated with CMOS and SOI technology to reduce cost. Results show the MEMS switch has excellent performance with insertion loss 0.3dB, return loss -27dB at 30GHz and high isolation -30dB at 40GHz. The process developed promises to simplify the design and fabrication of RF MEMS on silicon.

1. Introduction

As wireless communications has become an increasingly important part of everyday life, massive growth in the use of mobile phones and other radio communication systems has given rise to increasing demand for smaller, lighter and more reliable products with ever greater functionality. To achieve these performance requirements, novel design methodologies and an innovative approach to manufacturing technology are required. Capacitive switches are key system components in radio frequency microelectromechanical systems (RF MEMS) that provide superior performance, including low insertion loss, high isolation and zero static power consumption. The main challenge in this area is to integrate MEMS switches with CMOS, SOI technology at lower production cost than non-integrable devices and to simplify their design and fabrication.

Since the first MEMS switch appeared in 1979 [1], the technology has developed rapidly as evidenced by the number of research publications. Electrostatic capacitive RF MEMS shunt switches are one of the most popular switches, which offer superior performance such as low DC power consumption and high isolation. Referring to the 3D model Fig. 1, the bridge membrane is fixed both sides and straddles the CPW lines which are covered by a silicon nitride layer. Actuation voltage generates enough electrostatic force to pull the bridge down to contact the insulating layer. An RF signal can be passed through the CPW lines and reflected back since high capacitance is created. When DC bias is removed, mechanical energy stored in the bridge causes it to spring back to the freestanding position. The capacitance is dramatically reduced and produces high RF impedance. Once the switch is open, the RF signal can be transferred through the CPW lines.

A new switch design has been developed with an RF signal line width of 70µm, ground line width of 350µm and the distance between the two lines is 40µm. The characteristic impedance of CPW lines simulated by ADS software is 50Ω.

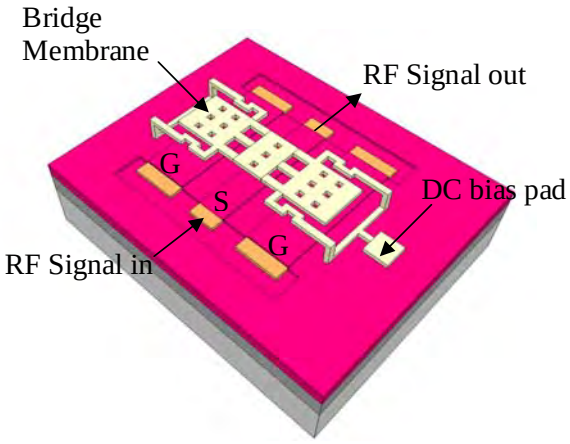


Figure 1. A 3D MEMS fixed – fixed beam switch over CPW lines.

2. Switch Design Considerations

A. Spring Constant

Referring to the 3D switch model shown in Fig. 1, the fixed-fixed beam membrane is supported off the substrate through four serpentine flexure springs. Such a design makes full use of the available space to enable the bridge membrane to move more flexibly, thus reducing the actuation voltage. If k_z is the z-direction spring constant for each of the springs, the total switch spring constant K_z is given by

$$K_z = 4\,k_z \tag{1}$$

$$k_z \approx \frac{48\,GJ}{l_a^2\left(\frac{GJ}{EI_x}l_a + l_b\right) * n^3}; \qquad n >> \frac{3\,l_b}{\frac{GJ}{EI_x}l_a + l_b} \tag{2}$$

[2]

For

$$J = \frac{1}{3}t^3 w\left(1 - \frac{192}{\pi^5}\frac{t}{w}\sum_{i=1,\,i\,odd}^{\infty}\frac{1}{i^5}\tanh\left(\frac{i\pi w}{2t}\right)\right)$$

As shown in formula (2)[2], the spring constant is dependent on the geometrical dimensions of the serpentine flexure, Young's modulus, the thickness of bridge membrane, the torsion constant and the number of serpentine flexures. The y-axis length of the beam and the number of serpentine flexures are the key parameters that affect the spring constant once the construction material is fixed.

Here, the Young's modulus of Al is 69GPa, Poisson's ratio is 0.31, shear modulus is 26.32GPa, x-axis moment of inertia is 13.33µm^4. As shown in Table I and Fig. 2, the serpentine flexure spring constant is calculated to be approximately 2.62N/m based on

formula (2) and the total switch spring constant is around 10.5N/m according to formula (1).

TABLE I. Serpentine Flexure Dimensions and Material Constant for the Shunt Switch.

Parameter	Calculation	Value
Thickness of Bridge Membrane t (μm)		2
X Axis Length of Beam l_b (μm)		20
Y Axis Length of Beam l_a (μm)		60
Width of Beam w (μm)		20
Number of Serpentines n		4
Al Young's Modulus E (GPa)		68.95
Al Poisson's ratio v		0.31
Shear Modulus G (GPa)	E/(2(1+v))	26.32
X-axis moment of inertia I_x(μm^4)	$wt^3/12$	13.33
Each Serpentine Flexure Spring Constant k_z (N/m)		2.62

B. Actuation (Pull-down) voltage

When electrostatic potential is applied between the top bridge and the bottom electrode, the electrostatic force pulls the top bridge down onto the insulating layer. The voltage at which the bridge deflects to approximately 2/3 of its original height is defined as the actuation voltage (V_a). As a rule of thumb, the breakdown voltage of the thin insulation layer should be approximately twice the actuation voltage.

$$V_a = \sqrt{\frac{8K_z g_0^3}{27A\varepsilon_0}} \tag{3}$$

Achieving voltages above 30V is quite a challenge for mobile phones, automotive vehicles, wireless devices and satellites, which rely on low voltage power supplies. Reduction of self-heating and the extension of device lifetime are also important considerations. Thus, reducing the actuation voltage not only broadens the range of possible applications, but also enhances switch reliability performance. According to formula (3), there are three main switch design considerations in the reduction of the actuation voltage: (i). Increase flexibility using spring design; (ii). Reduce gap height by trading off insertion loss; (iii). Increase coupling area within reasonable limits for miniaturized circuits.

The effective activating area is the sum of the two switch pull-down pad areas and centre pad area (143900 μm^2). The gap height g_0 between bridge membrane and CPW lines is 5μm. Thus, the actuation (pull-down) voltage is calculated as 17.5V.

C. On (Up) / Off (Down) state capacitance

The switch's pull-down pad and ground plane forms capacitor 1(C_1) and the centre area and signal plane works as capacitor 2 (C_2). Details can refer to Fig.3. The overall switch capacitance can be given as:

$$C = 2C_1 // C_2 \tag{4}$$

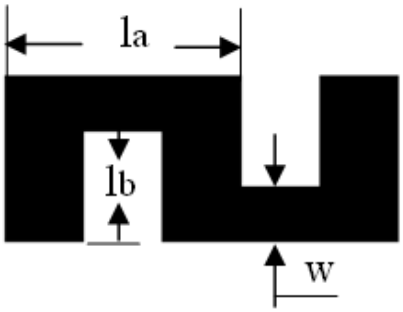


Figure 2. Schematic of a serpentine flexure spring.

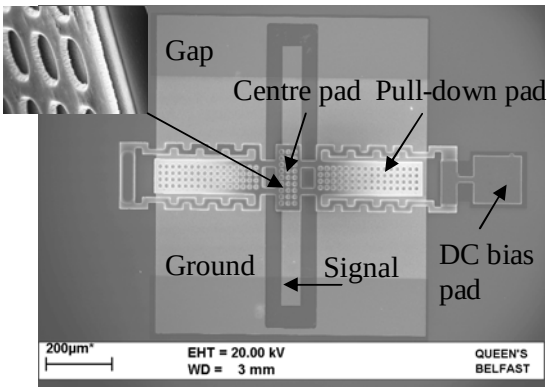


Figure 3. Top view of SEM picture of a MEMS switch

The on/off capacitance ratio is calculated as 152.5 by formula (5) according to the parameters in Table II.

$$\frac{C_{down}}{C_{up}} = \frac{\varepsilon_r g_0}{t_d} * \frac{2A_{1up} + A_{2up}}{2A_{1down} + A_{2down}} * \frac{A_{1down} A_{2down}}{A_{1up} A_{2up}} \quad (5)$$

TABLE II. Theoretical Parameters of the Switch.

Parameter	Value
Si_3N_4 Dielectric Constant $_r$	7.5
Down-state Contact Area A_{on} (μm^2)	114000
Up-state Contact Area A_{down} (μm^2)	143900
Down-state Capacitance C_{down} (fF)	3563.3
Up-state Capacitance C_{up} (fF)	23.4
On(Down)/Off(Up) Capacitance Ratio	152.5

It is tempting to make the dielectric layer as thin as possible in order to increase the capacitance ratio as shown in formula (5). However, it is impractical to deposit a Si_xN_y layer thinner than 200Angstroms due to pin-hole problems in thin dielectric layers. Also, the dielectric layer must be able to withstand the actuation voltage without dielectric breakdown. For this reason the layer is typically around 0.2µm thick in most switches built today.

3. Fabrication

The details of switch fabrication process shown in Fig. 4 are as the follows:

1) The switch is fabricated on an 8000Ω-cm high resistivity silicon wafer to reduce RF loss. Silicon dioxide and silicon nitride are grown and deposited for DC isolation. A 0.4µm tungsten layer is deposited by sputtering, patterned by mask 1 and etched to form the CPW lines Fig. 4(a).

2) A high quality insulator layer of low pressure chemical vapour deposition (LPCVD) Si_3N_4 (~0.2µm) [5] is deposited, patterned by mask 2 and etched by reactive ion etching (RIE). Thick photoresist is applied, patterned using mask 2 again and developed as the sacrificial layer to form anchor and contact pad Fig. 4(b).

Ex. 1016, Page 40

3) A thick layer of 2.0µm aluminium is sputter deposited and patterned to form the bridge. Finally, the sacrificial layer is removed by fuming nitric acid within 30 seconds and the switch is dried in IPA vapour Fig. 4(c). The device is held up-side-down to prevent stiction especially in water rinsing.

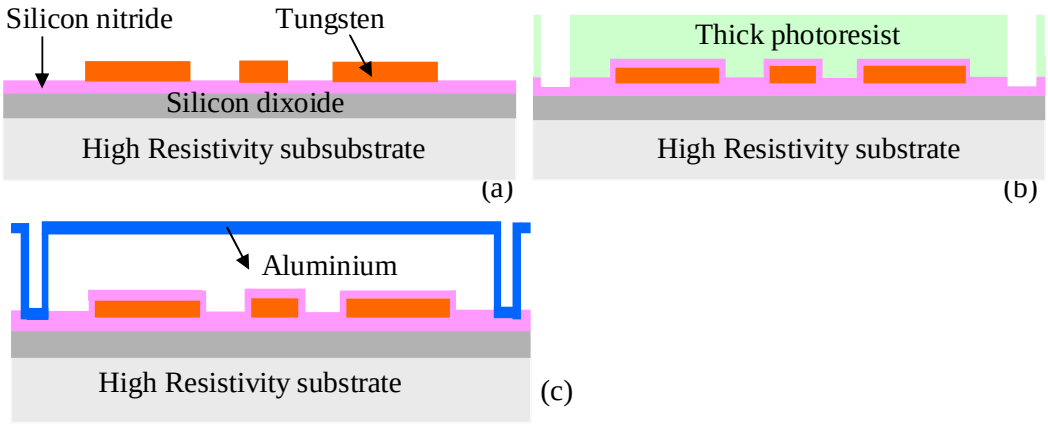


Figure 4. (a)-(c) Capacitive MEMS switch fabrication process

Comparing the process with recent publications [3] [4], there are four main advantages: (i). Only 3 masks required,; (ii). Tungsten used as CPW lines conductor instead of non-Si compatible material Au; (iii). High quality insulator layer; (iv). Fast fuming nitric release process instead of long traditional plasma release. Overall, the fabrication process is very fast and integrable with other electronic device on Si substrate.

4. RF Measurement Results

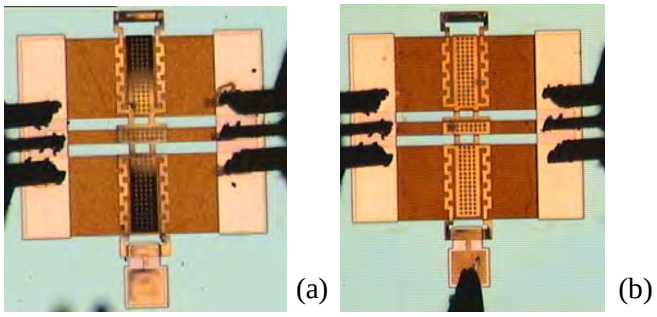


Figure 5. Optical image of a (a) On (up) -state (b) Off (down) - state switch

The optical image Fig. 5(a) under microscopy shows the MEMS switch in up-state with variation in colour, which indicates that the bridge is not totally flat at the same height due to residual stress and bridge own mass. On contrast, Fig. 5(b) the image illustrates a bright uniformly closed switch since bridge membrane contact CPW line completely.

The RF performance was characterized using an HP8510C vector network analyser. Actuation voltage is found to be 22V, which is in good agreement with the calculated value. The insertion loss of the tungsten CPW line without a switch is 0.95dB at 30GHz,

Ex. 1016, Page 41

Fig.6 (a). The insertion loss for CPW lines with an open state switch increases slightly to 1.3dB at 30GHz, Fig.6 (a). The switch itself therefore has a low insertion loss 0.3dB at 30GHz. Return loss is as low as -21dB at 40GHz. With the switch closed, isolation increases with frequency up to around -30dB at 40GHz, Fig.6 (b), due to the increased capacitive coupling.

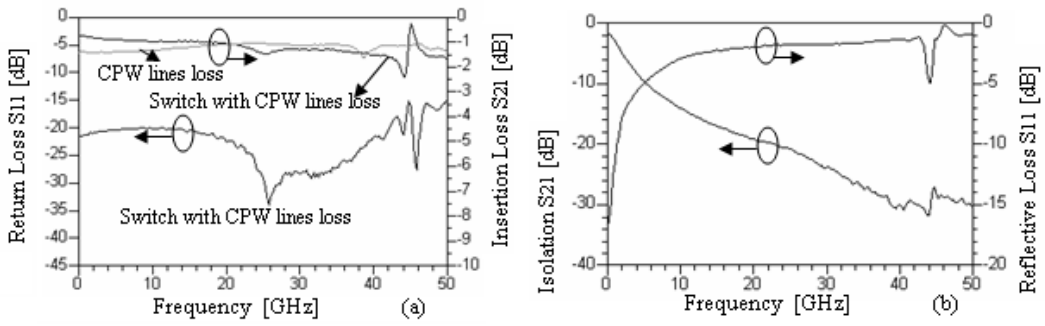


Figure 6. S-parameter measurement (a) CPW lines with & without open switch, (b) CPW lines with closed switch.

5. Conclusions

Electrostatic RF MEMS switches with a unique flexible spring were designed and fabricated using only 3 masks. The measurement results show the RF switches perform very well with low insertion loss and high isolation. Main fabrication advantages include integrable process with CMOS and SOI technology, a high quality LPCVD Si_3N_4 insulator layer and the fuming nitric acid fast release. This technology simplifies the switch fabrication process.

Acknowledgements

This work is sponsored by SoCaM project. The authors thank NISRC members and HF group staff at ECIT for their assistance.

References

1. K.E.Petersen, *IBM J. Res. Develop.*, **376**, 23 (1979).
2. G.M.Rebeiz, *RF MEMS: THEORY, DESIGN, AND TECHNOLOGY,* p.30, Wiley Interscience, New Jersey (2003).
3. R.Chan, R.Lesnick, D.Becher, Feng Milton, *Microelectromechanical Systems*, **713**, Journal of Volume 12, Issue 5 (2003).
4. A.Q.Liu, A.B.Yu, Q.X.Zhang, A.Alphones, H.M.Hosseini, *Sensors and Actuators A (Physical)*, **24**, 127 (2006).
5. A.Stoffel, A.Kovacs, W.Kronast, B.Muller, *J. Micromech. Microeng*, **1**, 6 (1996).

Ex. 1016, Page 42

ECS Transactions, 18 (1) 15-19 (2009)
10.1149/1.3096420 © The Electrochemical Society

CARBON CO-IMPLANTATION USED TO IMPROVE NMOS THRESHOLD VOLTAGE ROLL-OFF CHARACTERISTICS STUDY IN 65NM NODE CMOS TECHNOLOGY

Jinhua Liu[a], Allan Zhou, Jianhua Ju, Shugang Dai, ZhaoXu Shen, Eric Liu, Brisk Wang, Albert Hung, Hokmin Ho, Jay Ning

[a] Logic Technology Development Center, SMIC (BJ) Beijing 100176, PRC
Jinhua_liu@smics.com

Around 50mV RSCE and SCE improvement has been obtained by C co-implantation optimization in 65nm node CMOS technology. Besides this, gate to source/drain overlap capacitance (C_{gd0}) and junction capacitance (C_j) are seen to be reduced 6% and 11% respectively. C co-implantation before and after halo implantation are compared also, the result reveals that C co-implantation after halo implantation helps to improve NMOS drain induced barrier lowering (DIBL) features. Gate oxide leakage comparison for different C co-implantation conditions indicates that C co-implantation induces no gate oxide quality degradation.

INTRODUCTION

Device continuing scaling down the size according to Moore's law presents a challenge for threshold voltage roll-off control. Short channel effect (SCE) and reverse short channel effect (RSCE) are common phenomena in current devices, which are two main factors responsible for the anomalous threshold voltage roll-off trend along channel length. Though lots of mechanisms lead to these two effects, channel dopant diffusion, especially strengthened by transient enhanced diffusion (TED), is thought to be the most significant reason.

Recently, C co-implantation, used to reduce TED of boron has been much investigated (1-3). Substitution carbon in silicon is reported to capture the excess interstitials and suppress TED of ion-implanted interstitial-type diffusers in silicon.

In this paper, results of C co-implantation at the step of SDE implantation are comprehensively studied. Boron diffusion is evaluated by Secondary Ion Mass Spectroscopy (SIMS). NMOS with optimized C co-implantations are fabricated using 65nm technology. Device threshold voltage roll-off along gate length, driving current, gate to source/drain overlap capacitance, sub-threshold swing and junction leakage characteristics are demonstrated for the study of C co-implantation effects.

EXPERIMENT

Fig.1 gives the basic NMOS fabrication process flow used in the paper. Device goes through the STI isolation, gate stack, spacer, and source/drain implantation. Gate oxide thickness of the device is 21 angstrom (EOT). The lightly doped drain (LDD) is implanted with arsenic. Halo region is implanted with boron and indium, followed by tilt C co-implantation with the dose ranging from 3E14 to 8E14 and energy ranging from

Ex. 1016, Page 43

5KeV to 9KeV. Spike RTA is carried out for dopant activation at 1050ºC after deep source/drain implantation. NiSi is used as silicide in the process.

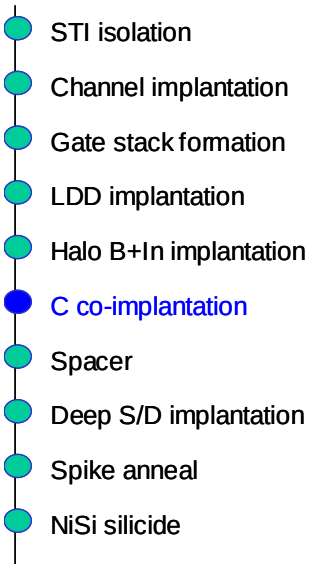


Fig.1 Process flow for NMOS fabrication

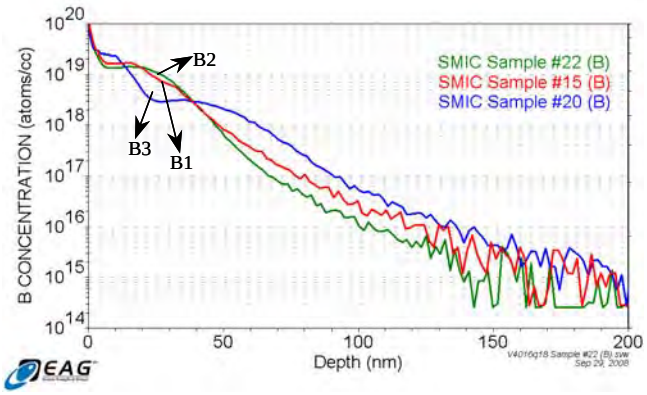


Fig.2 SIMS profile of boron with a dose of 4E13/cm^2. (B1) with C and after anneal; (B2) with C and without anneal; (B3) without C and after anneal.

SIMS analysis of boron profile with and without C co-implantation is shown in Fig.2. It is clear that boron profile with C co-implantation varies very little from the profile of boron as doped, while much different with the profile without C co-implantation. The result reveals that TED effect of boron is significantly suppressed by carbon as described in previous articles (4-5). This much suppressed boron diffusion improves device threshold voltage roll-off characteristics as shown in Fig.3. Both RSCE and SCE are decreased around 50mV. Besides this, junction capacitance (C_j) and gate to source/drain overlap capacitance (C_{gd0}) are also found to be much reduced as shown in Fig.4, around 6% for Cgd0 and 11% for Cj respectively. This lower Cgd0 may attribute to the higher boron concentration at the top surface of LDD region for device with C co-implantation as what is shown in Fig.2. And the lower boron concentration at the junction depth region (and deeper region) leads to a relatively lower Cj. This result is consistent with the data reported in previous paper (4).

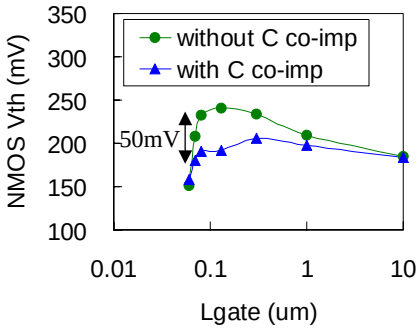


Fig.3. Threshold voltage gate length dependence

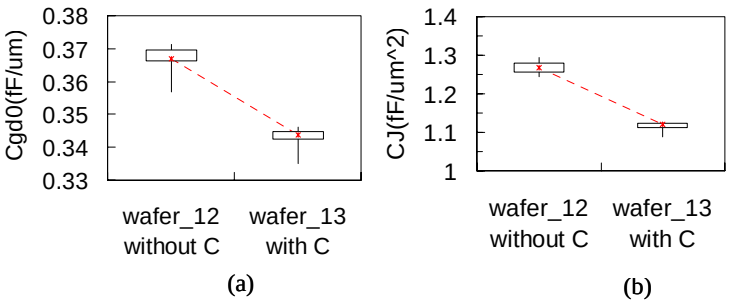


Fig.4 wafer with/without C co-implantation (a) Cgd0; (b) C_j

3.1. C co-implantation energy and dosage optimization

C co-implantation with different energy of 5KeV, 7KeV and 9KeV are carried out for comparison. From the threshold voltage roll-off figure, as shown in Fig.5, it is induced that C co-implantation energy comparable with halo boron implantation, which is 5KeV in this work, is most effective for SCE control, yet on the other side RSCE of this condition is seen to be the largest one among the splits. C co-implantation energy of 9KeV shows best RSCE but largest SCE. Though C co-implantation energy splits obviously impacts threshold voltage, Ion-Ioff performance of C co-implantation energy splits lead to nearly no difference as shown in Fig.6, which indicates that tilt implanted C does not degrade carrier mobility of the device. Fig.7 shows junction leakage comparison of the C implantation energy splits, no difference is seen for the median data, while higher energy condition exhibits relatively larger variation.

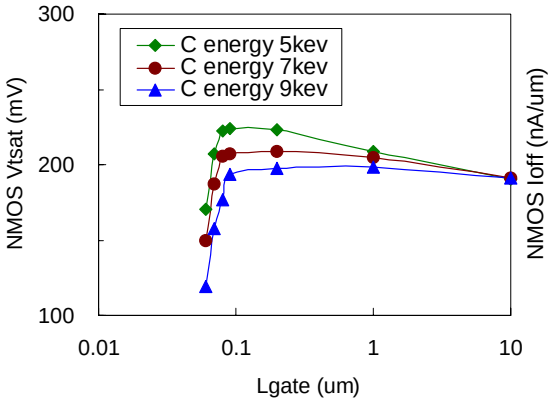


Fig.5 Threshold voltage roll-off characteristics for different C co-implantation energy.

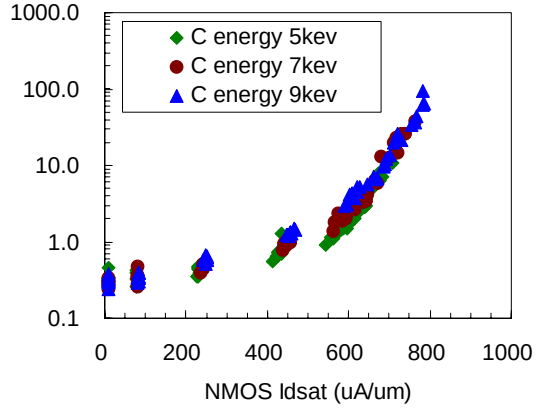


Fig.6 Ion-Ioff characteristics for different C co-implantation energy.

Three C co-implantation dose conditions, 3E14/cm^2, 5E14/cm^2 and 8E14/cm^2, are performed for analysis. As shown in Fig.8, C co-implantation with dose of 3E14/cm^2 shows 20mV larger RSCE than the other two conditions, while SCE of the three conditions displays no much difference. Compared with data as shown in Fig.5, we may conclude that device threshold voltage is much more dependent on C co-implantation energy than dose. Moreover, Ion-Ioff performance of C co- implantation dose splits shows no difference either, as shown in Fig.9. Junction leakage median data for different carbon dose exhibits no difference, but for 5E14/cm^2 case, data variation is largest among the splits (Fig.10).

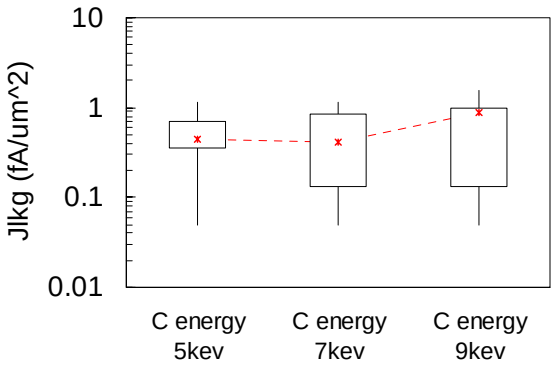


Fig.7 Junction leakage of different C co-implantation energy, higher energy shows larger variation.

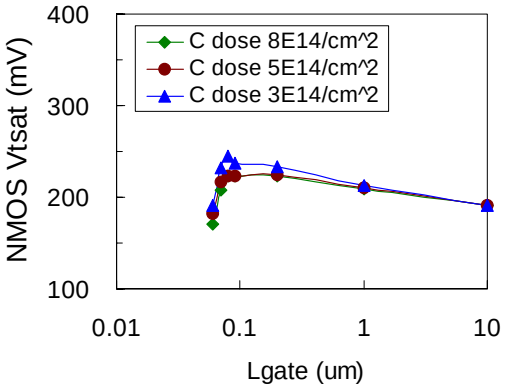


Fig.8 Threshold voltage roll-off characteristics for different C co-implantation dose.

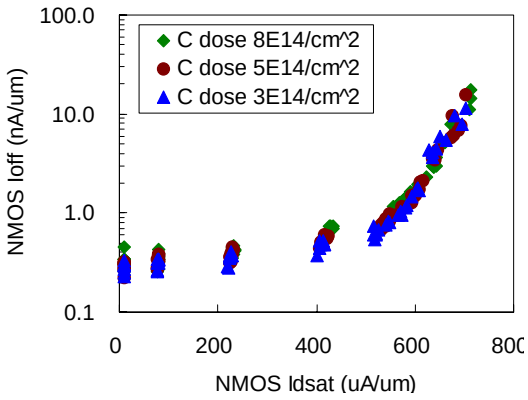


Fig.9 Ion-Ioff characteristics for different C co-implantation dose.

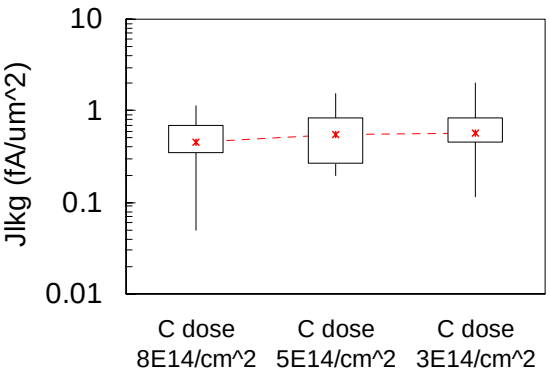


Fig.10 Junction leakage of different C co-implantation dose.

observed that when changing C co-implantation energy from 5KeV to 9KeV or changing C co-implantation dose from 3E14/cm^2 to 8E14/cm^2, Jg variation is just among 2nA/um^2 (5%), which indicates that tilt C co-implantation does not impact gate oxide robustness.

Considering device threshold voltage roll-off characteristics and Ion-Ioff performance as well as junction leakage together, C co-implantation with energy of 5KeV and dose of 8E14/cm2 is chosen as the best condition for the study of C co-implantation sequence.

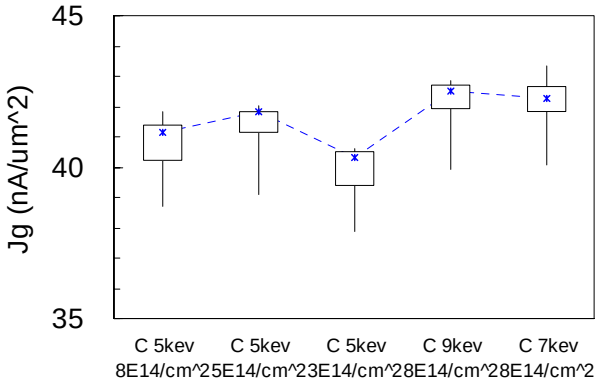


Fig.11 Gate oxide leakage of different C co-implantation conditions.

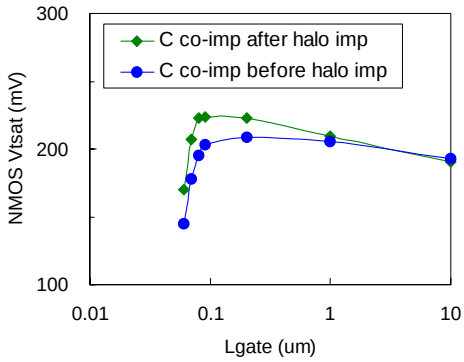


Fig.12 Threshold voltage roll-off characteristics for different C co-implantation sequence.

3.2. C co-implantation sequence: before and after halo implantation

C co-implantation before and after halo boron and indium implantation is studied to learn more about C co-implantation effect.

Threshold voltage roll-off characteristics of the splits, as shown in Fig.12, demonstrates that RSCE of C co-implantation after halo implantation is around 20mV larger than that of C co-implantation before halo implantation. However, drain induced barrier lowering (DIBL), data symbolized by threshold voltage difference of device at linear mode and saturation mode, as shown in Fig.13, shows that C co-implantation after halo implantation is around 8mV (6%) lower than the other condition, though Ion-Ioff of the two splits exhibits no difference as displayed in Fig.14.

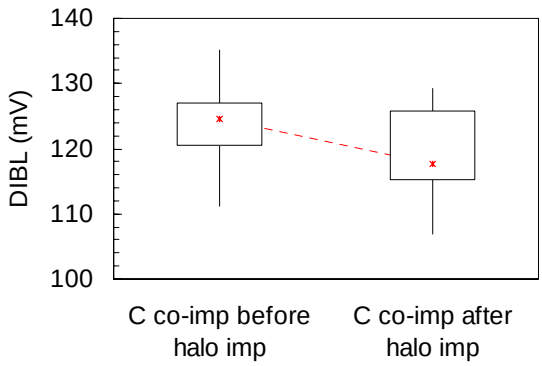


Fig.13. DIBL data for different C co-implantation sequence.

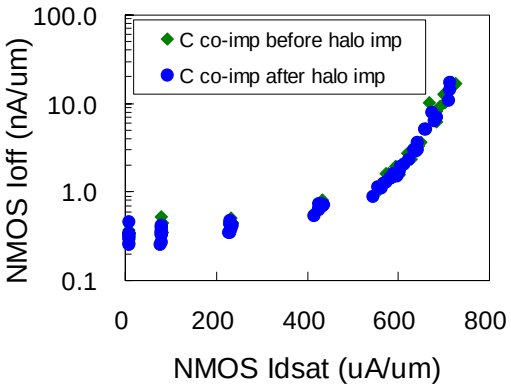


Fig.14 Ion-Ioff characteristics for different C co-implantation sequence.

CONCLUSION

By adopting tilt C co-implantation, both RSCE and SCE improvement in NMOS is obtained. C co-implantation also leads to Cgd0 and Cj reduction as much as 6% and 11% respectively. The experiment also reveals that C co-implantation with energy comparable with that of halo implantation, and dose to a level of 8E14/cm^2 is most effective for device threshold voltage roll-off control. This conclusion is generated by comprehensively considering of device Ion-Ioff performance and junction leakage features. It is also proved that C co-implantation induces no gate oxide quality degradation.

References

1. Ibrahim Ban, IEEE Trans. Electron. Dev. Vol.44, pp.1544-1551, 1997.
2. Akira Mineji and Seiichi Shishiguchi, Junciton Technology, IWJT'06, pp.84-87, 2006.
3. S.Gennaro, B.J. Sealy, Ion Implantation Technology. 2002, Proceedings of the 14th International conference, pp.552-555, 2002.
4. H.-J. Gossmann*, C. S. Rafferty, G. Hobler, H.-H. Vuong, D. C. Jacobson and M. Frei, Electron Device Meeting 1998. IEDM'98. pp.725-728, 1998..
5. Chung Foong Tan, VLSI Technology, Systems and Applications, 2008. VLSI-TSA2008, pp.32-33, 2008.

ECS Transactions, 18 (1) 21-26 (2009)
10.1149/1.3096421 © The Electrochemical Society

Investigation into the Output Characteristics and Improvement of Operation Margin of IMOS (Impact Ionization MOS) Devices

Huichu Liu, Ru Huang*, Zhenhua Wang
Department of Microelectronics, Peking University, Beijing 100871, China
*Email: huangr@ime.pku.edu.cn

Abstract-- The output characteristics of typical Impact Ionization MOS (IMOS) devices is comprehensively investigated in this paper. The results show that IMOS devices exhibit a narrow operating margin with non-saturated on-current. To improve the output characteristics, a new structure named "pocket IMOS" is proposed. Compared with the typical I-MOS devices, the "pocket I-MOS" can achieve a wider operating margin with lower source-drain bias and enhanced output performance.

Introduction

To achieve the continuous scaling down of CMOS technology, the power consumption issue has become one of the crucial limitations for integrated circuits application. Due to the non-scaling subthreshold slope SS of traditional MOSFETs, the high leakage current in the subthreshold region increases exponentially with the reduction of threshold voltage V_{th}, which results in significant static power consumption. Due to the fundamental limitation of kT/q in the Fermi-Dirac distribution of the source carriers, it is necessary to change the conventional carrier injection mechanism of diffusion in MOSFETs to realize the further reduction of SS down to 60mV/dec.

First reported by Gopalakrishnan *et al.*, the impact ionization metal-oxide-semiconductor (IMOS) devices has become one of the candidates for future ultra-low power application, which can achieve ultra-low (much lower than kT/q) SS at room temperature [1]-[2]. Based on a gated p-i-n diode, IMOS uses impact ionization as the carrier injection mechanism and the modulation of avalanche breakdown for on-off state switching. Recently, the operating voltage reduction approaches as well as the material engineering and the compatible fabrication technology applied in novel devices (e.g. double-gate, Fin-FETs and nanowires) have been investigated [3]-[7]. Simulations have also shown the influence of doping concentration and geometrical parameters of IMOS devices based on silicon-on-insulator (SOI) [8].

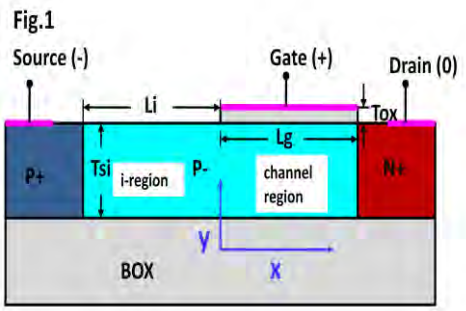
In this paper, we focus on the output characteristics of IMOS devices. In the first part, we will show the comparably narrow operating margin of typical IMOS devices, a detailed analysis of the non-saturation issues and the dynamic range limitations of on-current performance. In the second part, we will propose a new structure of IMOS devices by introducing a highly-doped "pocket-region" in the impact ionization region (i-region) of traditional IMOS transistors to improve the output characteristics.

Analysis on the Output Characteristics of IMOS Devices

The investigation is based on the typical structure of SOI n-type IMOS devices as shown in Figure 1. The source and drain regions are oppositely doped with the doping concentration of 1×10^{20} cm^{-3}. The channel region and i-region are low-doped regions with the same concentration (N_{sub}) of $1\times10^{15} cm^{-3}$ for practical cases. The thickness of the gate oxide layer and the silicon film is 4nm and 80nm, respectively. Both of the

Ex. 1016, Page 49

channel length and i-region length are 100nm. Simulations are performed with Synopsys T-CAD tools Sentaurus.



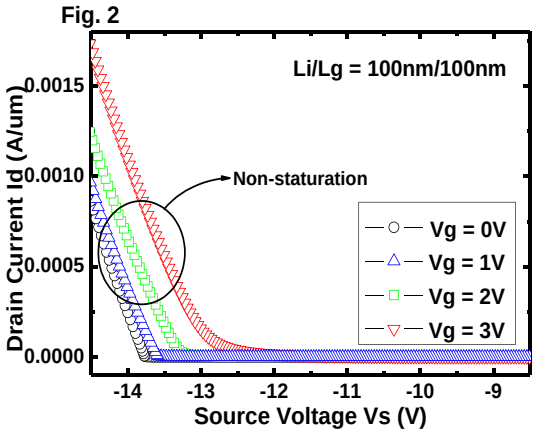


Figure 1. A schematic view of a typical n-IMOS transistor. The channel length (L_g), i-region length (L_i), the oxide thickness (T_{ox}) and the silicon film thickness (T_{si}) are 100nm, 100nm, 4nm and 80nm, respectively. Figure 2. Output curves of a typical n-IMOS transistor. The drain is zero biased and the gate voltage is varied from 0V to 3V.

A. Analysis of the Turning-on Behavior

Figure 2 shows the simulated typical output characteristics of n-IMOS devices. It can be observed that the drain current increases significantly with the V_s bias (V_d is grounded) under different gate voltages. Following we will analyze the non-saturation behavior.

Based on the impact ionization mechanism of carrier generation, the non-saturated on-current performance of IMOS devices can be explained by the dependence of impact ionization coefficients for electrons and holes on the effective electrical field, which can be given as:

$$\alpha_{electron} = \alpha_{\infty,electron} \bullet \exp[-(E_{crit,n} / E_{eff,n})^N] \quad (1)$$

$$\alpha_{hole} = \alpha_{\infty,hole} \bullet \exp[-(E_{crit,p} / E_{eff,p})^N] \quad (2)$$

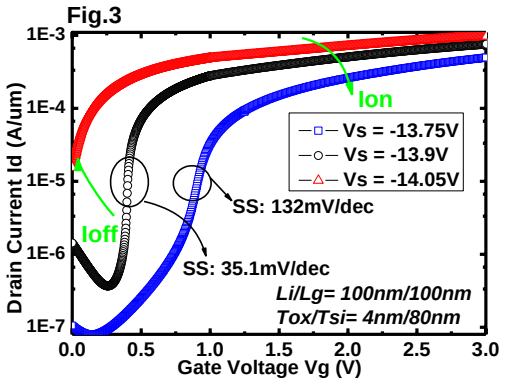
where $E_{eff,n}(E_{eff,p})$ and $E_{crit,n}(E_{crit,p})$ are the effective field for electrons (holes) and the critical electric field required to trigger the avalanche breakdown, respectively. $\alpha_{\infty,electron}$ and $\alpha_{\infty,hole}$ are asymptotic values which indicate the limit of impact ionization coefficient, and *N* varies between 1 ~2, depending on the specific model.

Due to the exponential relationship between $\alpha_{electron}$ and $E_{eff,n}$, the on-current of IMOS devices cannot reach saturation with the increase of the field in i-region [10]. From this point, the driving field is mainly affected by two factors. Firstly, with the modulation of the effective length of impact ionization region, the electric field needs to be controlled by the gate voltage V_g to induce the avalanche breakdown for switching. Secondly, as a high source-drain voltage is required for breakdown, the increase of V_{ds} can also lead to the increase of electrical field along the channel direction to generate more carriers for deeper breakdown. To enhance the saturation characteristics of the on-current in IMOS devices, the electric field should be effectively controlled by the gate voltage.

B. Investigation into the Operating Margin

The most significant advantage of IMOS is the ultra-low SS at room temperature. The hot carrier issues, which will induce a significant degradation of the device switching between on-off states, have been investigated [9]. However, according to our simulations, the bias of source-drain voltage is also one of the main factors which greatly affect the abruptness of switching between on-off states in IMOS devices and results in a comparably narrow operating margin of output characteristics.

The simulated transfer curves at different source-drain bias are illustrated in Figure 3. The typical IMOS devices can achieve the best SS of 35.1 mV/dec when V_s is biased at -13.9V. However, when the bias point increases or decreases, even the variation within a small range between -0.2V~+0.2V, the degradation of SS can be clearly observed. In addition, the on and off-current performance changes significantly with V_s. It indicates that the source-drain bias should be strictly controlled within a narrow operating margin, in order to maintain the superior switching characteristics for IMOS transistors.



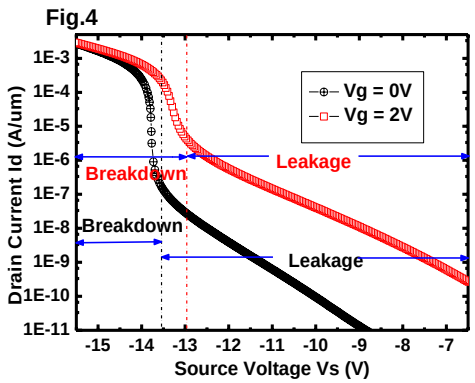


Figure 3. Transfer curves of typical n-IMOS. V_s varies from -13.75V to 14.05V. The best SS of 32.1 mV/dec is achieved at V_s=-13.9V. SS, I_{on} and I_{off} degrade with V_s variation. Figure 4. Output curves of IMOS devices at V_g=0 and V_{gs}=2V with log-axis. Body breakdown occurs at high $|V_s|$ bias.

This unanticipated phenomenon can be explained by the breakdown mechanism of p-i-n structure. As it is illustrated in Figure 4, avalanche breakdown in the body of p-i-n structure will be triggered as $|V_s|$ increases. To achieve the modulation of the length of i-region to trigger the avalanche breakdown, V_g should effectively control the electrical field distribution in the channel. In the ideal case, with the inversion layer forming at high gate voltage, the IMOS device can be seen as p-i-n diodes where L_i is the effective length for impact ionization. However, the impact ionization region cannot behave as the ideal case.

The electrical field and electrical potential distribution along the channel direction at 2nm away from the Si/SiO_2 interface during the on-state and off-state are given in Figure 5(a), respectively. Considering the surface field modulation by V_g, the breakdown voltage in the surface operation of IMOS devices is larger than the body p-i-n diodes [9]. Moreover, when the device is switched into on-state, the electrical field modulation induced by V_g is relatively ineffective to enhance the average of electrical field in i-region compared with the impact of high V_{ds} (above 10V). Therefore, it can be observed from Figure 5(a) that

the electrical field increases remarkably only at the edge of the gate (with a peak of 2.275×10^6V/cm), which cannot efficiently influence the impact ionization in 100nm i-region (with the lowest increasing point of 1.153×10^6V/cm). It can be also indicated by the electrical potential distribution. As shown in Figure 5(b), the increased electrical potential at the edge of i-region is 2.575V from off to on-state, which cannot effectively modulate the avalanche breakdown.

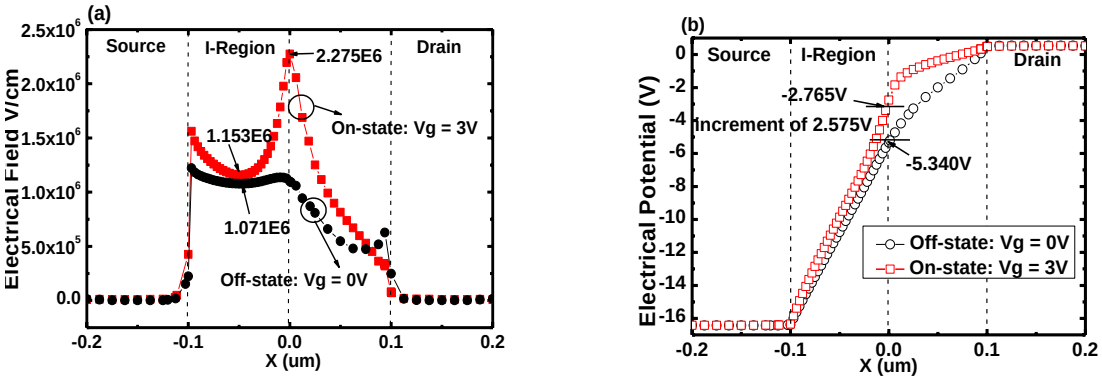


Figure 5(a). The field distribution in typical IMOS along the channel direction at 2nm away from the Si/SiO_2 interface during the on-state (V_g=3V) and off-state (V_g=0V) while V_{ds} is fixed at 16V.
Figure 5(b). The potential distribution along the channel direction at 2nm away from the Si/SiO_2 interface during the on-state (V_g=3V) and off-state (V_g=0V) while V_{ds} is fixed at 16V.

Due to the ineffective modulation of the electrical field in i-region, the required source-drain bias of IMOS devices will be very high and close to the breakdown voltage of body p-i-n structure. The increase of $|V_s|$ will induce the body breakdown with high leakage-current which cannot be controlled by V_g. In addition, as the avalanche breakdown current exponentially decreases with the decrease of the effective electrical field, the lower bias of V_s will significantly reduce I_{on}, which can also impact the ratio of I_{on}/I_{off}. Thus, the typical IMOS devices show a narrow operating margin which might limit the device practical application.

New Device Structure with Improved Output Characteristics

A. Device Structure

To improve the electric field distribution in the on-state and realize a wider operating source-voltage tolerance, a new structure of I-MOS devices is proposed by introducing a pocket-region in i-region in this paper. The structure is schematically shown in Figure 6(a). The “pocket region” is a very narrow highly n-doping region in the n-IMOS.

With the newly-introduced pocket region as shown in Figure 6(b), two p-n junctions along the channel direction can be formed to induce peak electrical field at the edge of the gate while enhancing the average effective field for avalanche breakdown. Along the channel in x direction, the two p-n junctions are biased backward and forward with the supply V_{ds}, respectively. As shown in Figure 7(a), at the on-state, the enhanced electrical field in i-region ranging from 1.78×10^6 to 2.33×10^6 V/cm will enhance the impact ionization coefficients. Moreover, the inversion layer formed by the positive V_g will extend along the channel to the “pocket region” to achieve a continuous carrier

Ex. 1016, Page 52

distribution, which will further improve the on-current performance. Additionally, to achieve low leakage, the pocket region should be depleted at the off-state, which can be realized by minimizing the length of pocket region $L_{i,poc}$ (10nm or less for 100nm L_i). Furthermore, the "pocket region" in IMOS devices will also redistribute the potential along the channel direction as shown in Figure 7(b), with an increment of 5.281V distributed in i-region from off to on-state.

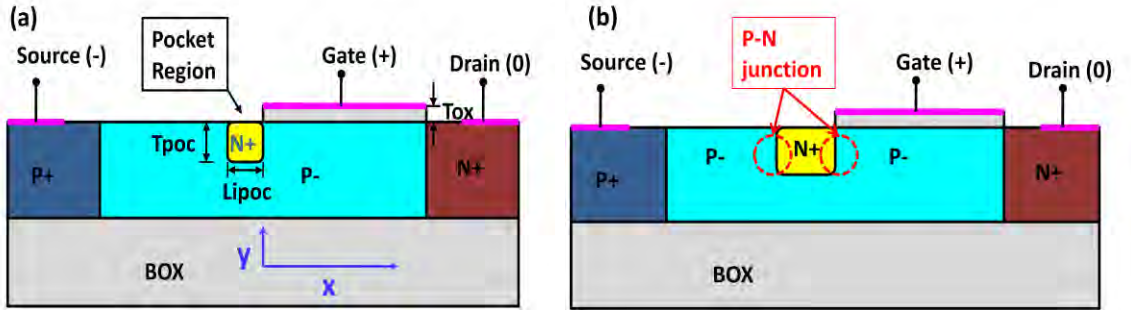


Figure 6(a). A schematic view of a pocket n-IMOS transistor. The doping concentration in the pocket-region is $1 \times 10^{19} cm^{-3}$. Both Tpoc and Lipoc in the simulated pocket IMOS are 10nm.
Figure 6(b). With V_{ds} biased postively, two p-n junctions will form to affect the field distribution.

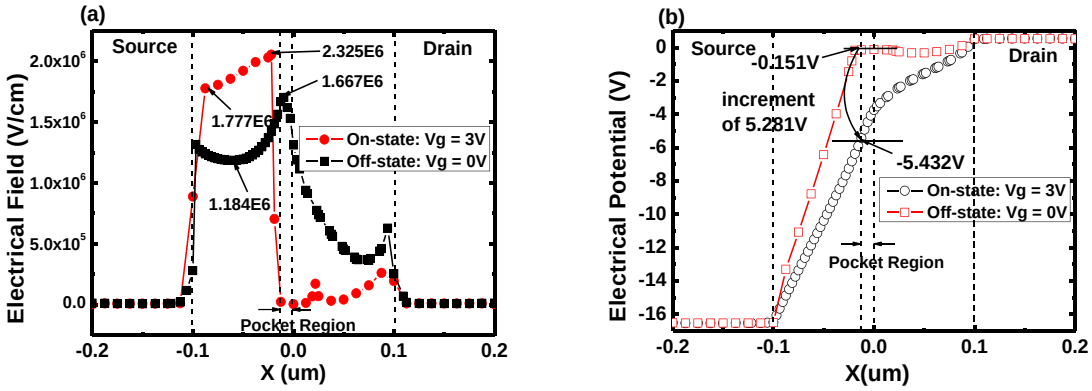


Figure 7(a). The field distribution in pocket IMOS at on (Vg=3V) and off-state (Vg=0V) along the channel direction to 2nm of the Si/SiO2 interface while Vds is fixed at 16V. Figure 7(b). The potential distribution in pocket IMOS along the channel direction to 2nm of the Si/SiO2 interface.

B. The Improvement on Output Characteristics of the Pocket IMOS

In Figure 8(a), the output characteristics of "pocket n-IMOS" are presented with the thickness of pocket region T_{poc} and $L_{i,poc}$ of 10nm. Compared with the traditional n-IMOS, the pocket n-IMOS shows a wider operating margin at lower source-drain bias and enhanced saturation characteristics. The dynamic range can be defined for low leakage current less than 1uA/um and breakdown current larger than 1mA/um. The V_s of pocket IMOS can achieve an operating margin of more than 800mV from -13.4V to -12.6V, while the V_s of traditional IMOS devices should be within the range of 300mV. In addition, with the enhancement on electrical field in i-region during on-state, the source-drain voltage limitation to trigger the avalanche breakdown is effectively reduced to 12.6V or even less in pocket IMOS compared with -13.7V in traditional IMOS devices. As a result of the improved gate control in the electrical field, the on-current can be increased while the output conductance can be reduced into 0.8916ms at $V_s = 2V$ compared with 1.1323ms in traditional IMOS devices as illustrated in Figure 8(b). And

Ex. 1016, Page 53

the superiority of ultra-low SS is still maintained in "pocket IMOS" within the operating margin. Further study on the impact of geometrical parameters and detailed analysis of physics principles of the pocket region will be reported in the future.

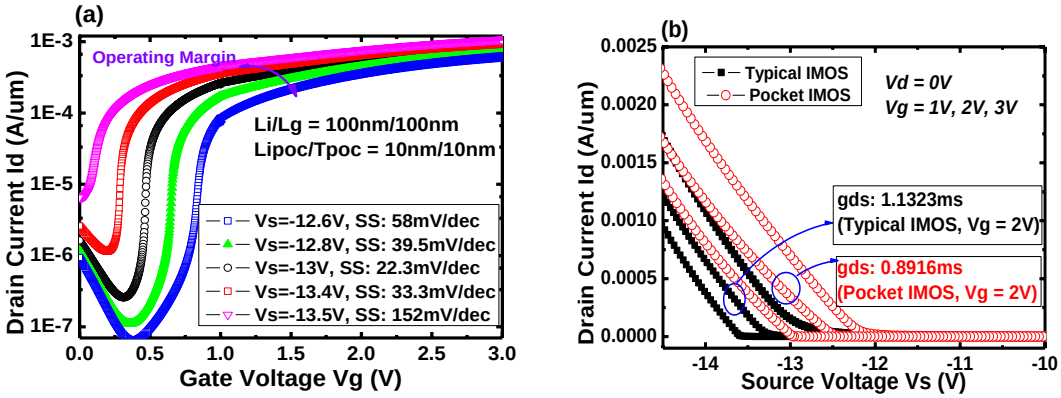


Figure 8(a). Transfer curves of pocket n-IMOS. L_g, L_i, T_{ox}, T_{si}, L_{ipoc} and T_{poc} are 100nm, 100nm, 4nm, 80nm,10nm and10nm, respectively.
Figure 8(b). Output curves of pocket IMOS and traditional MOSFETs at Vg=1V, 2V, 3V.

Conclusion

In this paper, the non-saturation issues and operating margin of typical IMOS devices have been investigated. IMOS shows a narrow operating margin within 300 mV for the source-drain bias and non-saturate on-current which is greatly impacted by the bias point. The main reason lies in the effectiveness of gate control on the electrical field distribution in the impact ionization region. To improve the operating margin and achieve more robust devices, a new structure of IMOS devices, "pocket IMOS", has been proposed by inserting a highly-doped pocket region in i-region close to the gate edge. The "pocket IMOS" performs an enhanced operating margin of 2.8 times compared with the traditional device structure as well as lowered V_s bias and enhanced on-current characteristics with the subthreshold slope still maintained ultra-steep. The improvement by "pocket IMOS" will achieve the stable operating of IMOS devices for practical application.

Acknowledgement

This work was supported in part by the NSFC(60625403), 973 Projects(2006CB302701), and the NCET Program.

References

1. K. Gopalakrishnan et al., *IEDM*, p. 289−292, (2002).
2. W. Y. Choi et al., *IEEE Trans. Nanotechnol.*, p. 322−325, (2005).
3. W. Y. Choi et al., *IEDM*, p. 203−206, (2004).
4. E.-H. Toh et al., *IEDM*, p. 971-974, (2005).
5. E.-H. Toh et al., *IEEE Electron Device Lett.*, p. 975–977, (2006).
6. E.-H. Toh et al., *IEDM.* p. 195-198, (2007).
7. F.A. Hassani et al., *AFRCON.* p. 1-7, (2007).
8. F. Mayer et al., *IEEE Trans. Electron Devices*, p.1852-1857, (2006).
9. K. Gopalakrishnan et al., *IEEE Trans. Electron Devices*, p. 77–84, (2005).
10. Z. Wang et al., ICSICT, p. 72-25, (2008).

Ex. 1016, Page 54

ECS Transactions, 18 (1) 27-31 (2009)
10.1149/1.3096422 © The Electrochemical Society

Minimizing device variation caused by RTA temperature variation across the shot field

Jianhua Ju*, Eric Liu, Shugang Dai, Allan Zhou, Zhaoxu Shen, Jinhua Liu, Brisk Wang, Daniel Deng, Julie Tang, Albert Hung, Hokmin Ho and Jay Ning

Logic Technology Development Center, SMIC (BJ) Beijing 100176, PRC
*E-mail: Jianhua_Ju@smics.com

Spike Rapid Thermal Annealing (RTA) is a commonly used method for source and drain dopant activation in 65nm CMOS technology. It is found that the Idsat variation is very sensitive to RTA temperature. To better control the Idsat uniformity across the shot, a systematic study on STI and poly pattern density distribution across the shot field and their effects on the RTA temperature was carried out. It was demonstrated that the STI and poly pattern density variation across the field will result in more than 10ºC RTA temperature variation. To minimize the effect, an optimization of RTA ramp rate and peak temperature was studied. Source/drain (S/D) extension and S/D implant optimization has also been done for the reduction of PMOS device Idsat temperature sensitivity

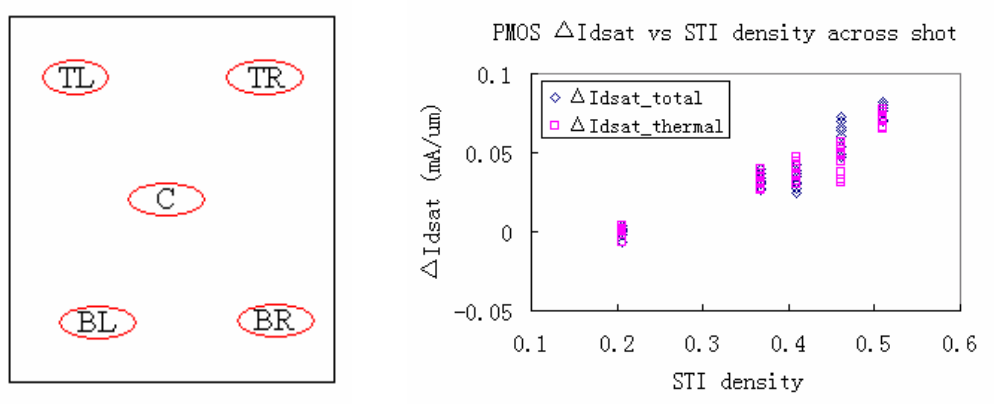
Introduction

It was reported that in 65nm CMOS processes, MOSFET transistor characteristics, e.g. Idsat and threshold voltage Vt are subject to the AA pattern density variation across the shot field (1). The device variation was caused by the RTA temperature difference due to AA pattern distribution. The variation is particularly pronounced for PMOS transistors. The active dopants for PMOS transistor are boron and BF2 speices. Both of them have a high diffusion coefficient and are more sensitive to the temperature variation as compared to other speices, e.g. phosphorous. Optimization of PMOSFET transistor RTA process by selecting proper peak temperature and temperature ramp up rate could improve within shot PMOS device uniformity (2). In this paper, a study of minimizing PMOS temperature sensitivity by means of optimizing source-drain extension and deep source-drain implantation dosage and improving within shot uniformity by adjusting RTA ramp rate and peak temperature were carried out.

Experimental Detail

The wafers used in this work were produced by a 65nm node low power logic technology. Process flow includes poly gate offset spacer and a high tensile contact etch stop layer to overcome SCE and improve NMOS devices performance. The samples are 300 mm P type, (100) epi-silicon wafers with resistance between 8-12 ohm/cm. Device test structures were placed five different sites within a shot field as shown in Fig. 1 (a). The Idsat variations of PMOSFET transistors across the shot as shown in Fig. 1 (b) with different S/D RTA ramp up rate, peak temperature and implant condition have been

measured. The detailed experiment conditions are listed in table 1. Idsat variations with STI and poly pattern densities at the five locations across shot are compared.



(a) (b)

Fig.1(a) The same test key locations across shot; (b) Idsat variation

Table 1 Sample ID and Experiment Condition

Sample ID	PMOS IMP condition	S/D RTA Condition	
		Peak temperature (C)	Ramp up rate (C/s)
HR_RTA	BL		Higher
MR_RTA	BL	Median	Median
LR_RTA	BL		Lower
HT_RTA	BL	Higher	
MT_RTA	BL	Median	Higher
LT_RTA	BL	Lower	
S/D extension Dose	PLDD dose -11%	Median	Lower
S/D Dose	S/D dose - 10%		

The poly CD at five locations within the shot has been monitored and STI effective density has been calculated as shown in table 2. The poly CD variation is due to poly etch loading effect, since poly density at five locations is different. The STI pattern density and poly transmission rate are also listed in table2, which are collected from the mask of the 65nm low power chip. The STI effective pattern density is defined as STI pattern density multiplied by poly transmission rate, which is proportional to localization temperature during thermal process (1). The total MOSFET transistor drain current (Idsat) variation at the five locations mainly comes from three parts: the thermal budget variation of source/drian (S/D) RTA process, the poly CD variation due to etch loading effect and the random variation of implant process, so that the total Idsat variation can be expressed as:

$$\Delta Idsat\_total = \Delta Idsat\_thermal + \Delta Idsat\_cd + \Delta Idsat\_random \quad [1]$$

ΔIdsat_thermal and ΔIdsat_cd in equation [1] refer to Idsat variations induced by thermal and poly CD variation respectively. ΔIdsat_random item can be ignored, as it has been observed that at five different location the implant process leads to nearly no variation.

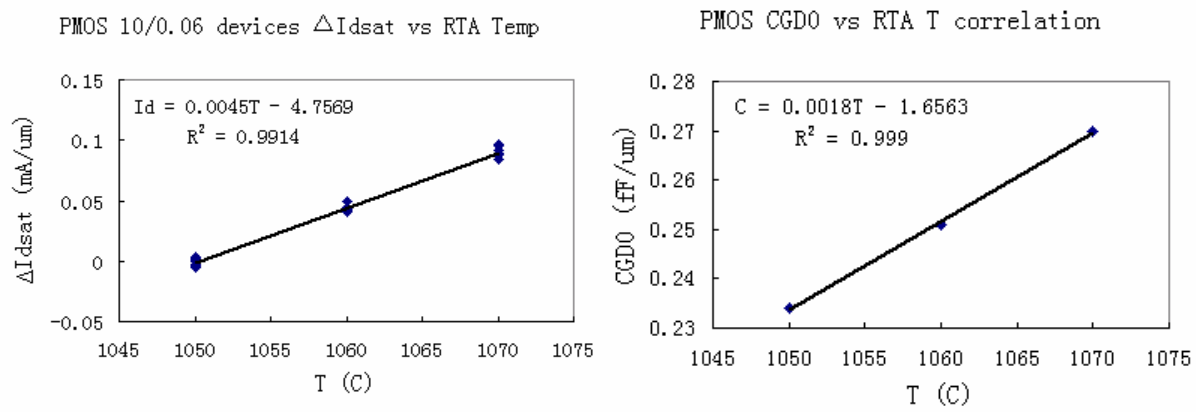
So that after removing ΔIdsat_random, it is easy to get from equation [1], ΔIdsat_thermal caused by STI effective pattern density from the expression:

$$\Delta Idsat\_thermal = \Delta Idsat\_total - \Delta Idsat\_cd \quad [2]$$

Table 2 Pattern density information and inline poly CD

Location within shot	TL	TR	BL	BR	C
Inline poly CD (nm)	64.9	63	62	59.1	65.1
STI pattern density (%)	54.9	45.63	61.48	64.96	58.21
Poly transmission rate (%)	66.75	45.7	82.92	71.03	70.2
STI effective density (%)	36.65	20.85	58.98	46.14	40.86

Equation [2] describes the Idsat variation caused by thermal budget difference, which is a function of the STI effective pattern density as shown in Fig.1(b). There is a nearly linear correlation between ΔIdsat_thermal and STI effective pattern density.



(a) (b)

Fig.2 PMOSFET's Idsat (a) and CGD0 (b) variation at different source and drain RTA peak temperature

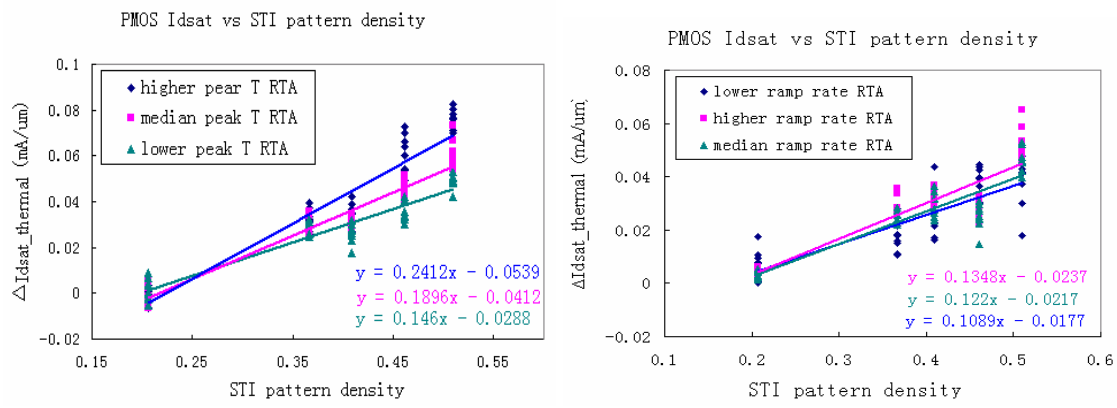
The correlation of Idsat variation with different S/D RTA peak temperature as shown in Fig.2 also shows a linear trend, which indicates STI effective pattern density and local temperature at five location across shot during the RTA process may have a very close (or even linear) correlation. Reducing the sensitivity of the ΔIdsat_thermal to STI pattern density means improvement of Idsat uniformity across the shot. Different experimental conditions shown in table 1 have been studied for optimization.

Result and Discussion

Fig.3 shows the relationship between Idsat variation and the STI pattern density at different source-drain RTA peak temperature. Comparing Fig.3 (a) and Fig.2 (a) data, we can find that the Idsat variation caused by STI pattern density difference within shot is more than 10ºC temperature variation. It is also observed from fig.3 (a) that low temperature RTA has low sensitivity of Idsat to STI pattern density. However, reducing the RTA peak temperature, PMOSFET Idsat will be degraded dramatically as shown in Fig.4 (a) due to lower boron activation (3).

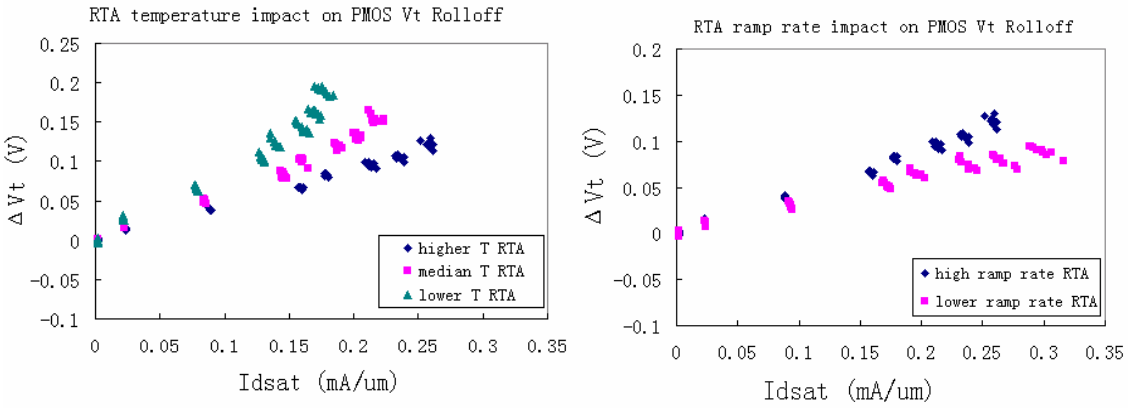
Ex. 1016, Page 57

Fig.3(b) shows the relationship between Idsat variation and the STI pattern density at different ramp up rate RTA process. Comparing the data (diamond and square), it is clear that lower ramp up rate RTA has lower sensitivity of Idsat to STI pattern density. Yet lower ramp rate RTA condition also leads to PMOSFET worse in short channel effect (SCE) as shown in Fig.4 (b). The ΔVt in the Fig.4 is the difference of threshold voltage between each channel length and channel length of 10um. This worse SCE mainly attributes to enhance boron diffusion caused by higher thermal budget of lower ramp rate RTA process. So that optimization of ramp up rate and peak temperature of S/D RTA is necessary to balance devices performance and Idsat uniformity across shot



(a) (b)

Fig3. The Idsat variation with STI pattern density at different RTA peak temperature (a) and ramp up rate (b)



(a) (b)

Fig.4 PMOSFET's Vt vs Idsat roll off characteristics at different RTA peak temperature (a) and ramp up rate (b)

Fig.5 shows the relationship between Idsat variation and STI effective pattern density with different S/D extension and deep S/D implant conditions. Comparing the data of diamonds and squares in Fig.5, it is clear that the sensitivity of Idsat variation to pattern density can be improved by reducing S/D extension dose. Reducing deep S/D implant dose can obtain the same result also. As Idsat variation to RTA temperature is nearly linear as what is shown in Fig.2 (a), we may conclude from Fig.5 that S/D extension and deep S/D implant dose optimization helps to improve Idsat to a RTA temperature sensitivity 13.5% and 8.9% respectively.

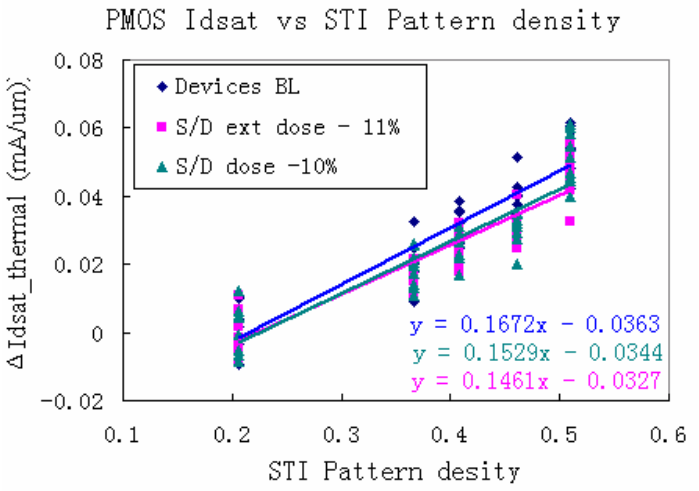


Fig.5 The Idsat variations to STI pattern density for different S/D extension and S/D implant conditions

Conclusion

To better control the Idsat variation across the shot, a systematic study on STI and poly pattern density distribution across the shot field and their effects on the RTA temperature was carried out. Experimental results show that lower RTA ramp rate and peak temperature will have low Idsat temperature sensitivity. However optimization of proper ramp rate and peak temperature is necessary. By reducing PMOS S/D extension and deep S/D implant dose it is proven that it is possible to reduce the PMOS Idsat temperature sensitivity to 13.5% and 8.9%, respectively.

References

1. I. Ahsan, et al., *2006 Symposium on VLSI Technology, Digest of Technical Papers*, pp. 170–171 (2006),
2. Kelin J. Kuhn et al., *IEEE International Electron Devices Meeting, IEDM Technical Digest*, pp. 471–474 (2007),
3. Al-Bayati et al., *In Proceedings of the 14th International Conference on Ion Implantation Technology*, pp. 185–188 (2002)

ECS Transactions, 18 (1) 33-37 (2009)
10.1149/1.3096423 © The Electrochemical Society

ELECTRICAL CHARACTERIZATION OF MOS MEMORY DEVICES WITH SELF-ASSEMBLED TUNGSTEN NANO-DOTS DISPERSED IN SILICON NITRIDE

Y. Pei [a], C. Yin [b], M. Nishijima[c], T. Kojima[d], H. Nohira[e], T. Fukushima[b], T. Tanaka[b], and M. Koyanagi[b]

[a]International Advanced Research and Education Organization, Tohoku University, Sendai, Japan
[b]Dept. of Bioengineering and Robotics, Tohoku University, Sendai, Japan
[c]Institute for Materials Research, Tohoku University, Sendai, Japan
[d]Technical Division, School of Engineering, Tohoku University, Sendai, Japan
[e]Musashi Institute of Technology, Tokyo, Japan

Crystallized W-NDs dispersed in silicon nitride were prepared by self-assembled nanodot deposition (SAND) with extremely high density ($1.3x10^{13}/cm^2$) and small size (1~1.5nm), successfully. XPS results show that the high metallic states of tungsten are remained after 800℃ anneal due to the surround silicon nitride. MOS memory capacitor with W-NDs floating gate and HfO_2 blocking dielectric was fabricated. A hysteresis memory window of ~ 1.0V is observed under low sweeping gate voltage of ±5V. A maximum memory window as large as 29V was obtained with sweeping gate voltage of ±23V, suggesting that it is a good candidate for multi-bit nonvolatile memory application.

Conventional flash memories have remained in the lead in the nonvolatile memory (NVM) market segment. In the near future, however, they will face some serious problems concerning the poor reliability of charge retention, endurance, and high power consumption with scaling down. Floating gate flash memory devices, which contain Si or Ge nanodots, have already attracted the attention of researchers as the promising candidate for future NVM [1]. It is expected that one can obtain longer retention time than in the case of using conventional floating gate memory devices. In such memories, each nanodot is electrically isolated from the rest, so that when there is charge leakage towards the substrate, this will involve only one or a few nanodots, but not the entire floating gate. This leads to high reliability. Additionally, these memories, owing to the fact that they operate with a small number of carriers, have low power consumption, very fast write-erase operation, and very high memory cell density. In recent years, metal nanodots floating gate flash memoris have attracted much attention for the future NVM device [2-4]. Compared with conventional semiconductor nanodots floating gate, the metal nanodots floating gate shows a lot of advantages such as higher density of energy state around Fermi level, wide range of available work functions. However, they will face some challenges such as prevention of metal/oxide reaction and metal diffusion into tunneling and blocking dielectrics during fabrication process, which may potentially degrade the device performance [5-6]. Another important challenge is fabrication of high density nanodots to overcome the fluctuation of electrical characteristics. It had pointed

Ex. 1016, Page 61

out that the main critical issue which limits the scaling of nanodots memory is fluctuation of the electrical characteristics [7]. This is due to the spread in dot size and dot density from one device to another. To overcome the electrical fluctuation each memory cells, the high density nanodots were required.

To improve the electrical performance of flash memory devices, efforts are also made to replace the SiO_2 blocking oxide with high dielectric constant materials. This improves controllability of gate to write/erase data, resulting in lower programming voltages [8].

In this work, to solve above issues, the metal–oxide-semiconductor (MOS) memory capacitor with a floating gate of high density self-assembled tungsten nanodots (W-NDs) dispersed in silicon nitride and HfO_2 blocking dielectric was fabricated. The silicon nitride is a good isolator to prevent the metal/oxide reaction as reported in reference [9]. The physical and electrical properties were discussed in detail.

Experiment

The MOS memory capacitor with W-NDs in silicon nitride was fabricated as the following process. p-type Si (100) substrates with a resistivity of 5~10 Ωcm were used. A 10-nm-thick W-NDs layer was deposited on 5-nm-thick tunneling SiO_2 by SAND method [5, 9]. In this method, tungsten chips placed on a Si_3N_4 target were cosputtered in high-vacuum RF sputtering equipment. The post deposition annealing (PDA) of W-NDs film was performed at 800℃ in high vacuum ambient. The blocking oxide layer is 40-nm-thick HfO_2 formed by RF sputtering equipment, followed by 600℃ annealing in nitrogen ambient. Finally, Al electrode with a diameter of 0.3 mm was evaporated. As reference, in this study, the MOS capacitor with silicon nitride layer without W-NDs was fabricated simultaneously. Figure 1 shows the schematic cross-sectional view of a W-NDs memory capacitor. The physical characteristics of W-NDs were evaluated by high-resolution transmission electron microscopy (HRTEM) and x-ray photoelectron spectra (XPS). The memory properties were investigated using an Agilent B1500A semiconductor parameter analyzer at room temperature.

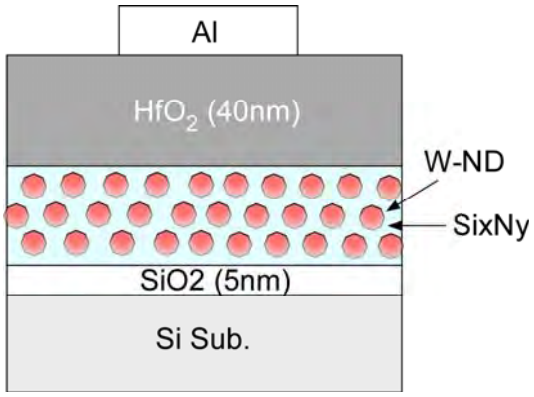


Figure 1. Schematic cross-sectional view of W-NDs MOS memory capacitor.

Results and Discussion

A HRTEM image of W-NDs with 800°C PDA was shown in Fig. 2. It is clearly observed that the W-NDs were dispersed in silicon nitride film with a diameter of 1~ 1.5nm. The density of W-NDs was estimated at ~$1.3x10^{13}/cm^2$, which is higher than another group by one or more orders [3-4]. Crystallized W-ND with visible lattice fringes was shown obviously in the inset of Fig. 2. These results indicate that the W-NDs are

Ex. 1016, Page 62

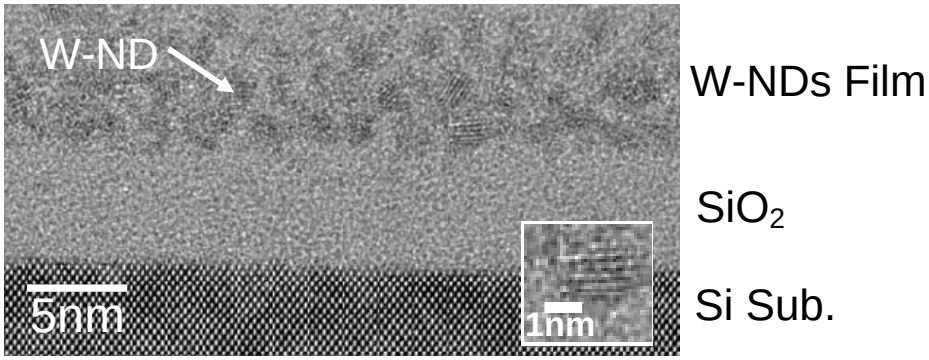


Figure 2. HRTEM cross-sectional image of W-NDs film with 800°C PDA. The W-NDs size is 1~1.5nm, density is ~$1.3x10^{13}/cm^2$.

nano-crystallized. The excellent interface at SiO_2/W-NDs is observed, indicating that the W-NDs are not diffused into the tunneling oxide even after 800℃ anneal.

To confirm the chemical composition of W-NDs after 800℃ PDA, the XPS spectrum of W4f was shown in Fig. 3. In a previous study, a tungsten film showed a spin-orbit doublet due to W $4f^{7/2}$ at 31.4 eV and W $4f^{5/2}$ at 33.6 eV, which represent W^0 [10]. With processing of metal/oxide reaction, W4f peaks were shifted to high binding energy. From Fig. 3, we found that the W-NDs have high metallic states and partly oxidized composition, which is identical to the high binding energy. Compared with W-ND embedded in silicon oxide reported in ref. [5], the oxidation of tungsten was suppressed by silicon nitride markedly.

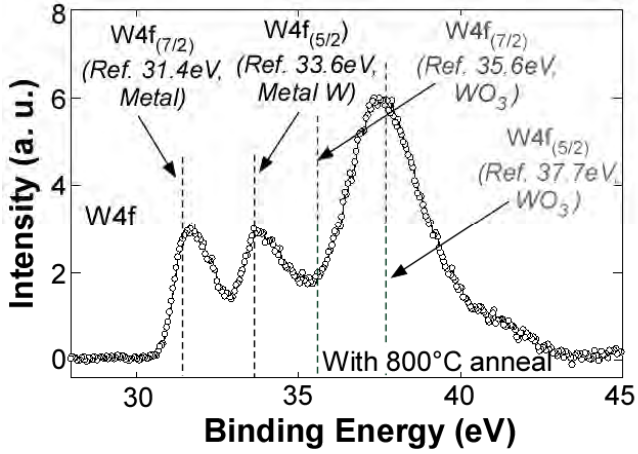


Figure 3. W4f level XPS spectrum of W-NDs in silicon nitride with 800°C PDA.

In Fig. 4, we show the capacitance-voltage (C-V) characteristics of W-NDs MOS memory capacitor, measured at different frequencies with a range of 1KHz ~1MHz. The frequency independence of C-V curves suggested that the memory window is not due to the interface traps. Figure 5 shows the C-V measurement with bi-directional voltage sweeping from 10 to –10V and from –10 to 10V for MOS memory capacitors with and without W-NDs. Considering the small C-V hysteresis of MOS capacitor with silicon nitride without W-NDs, the larger counterclockwise hysteresis of 7 V for W-NDs MOS

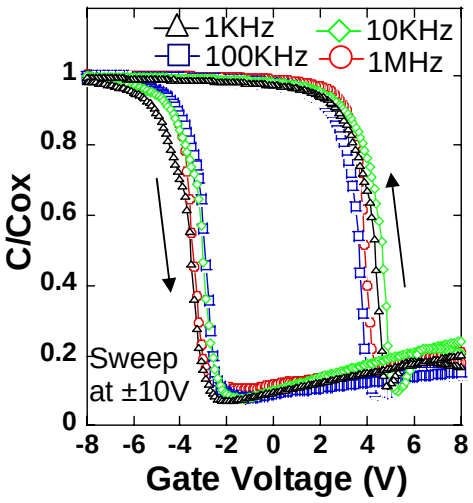


Figure 4. C-V curves of W-NDs MOS memory capacitor measured at different frequencies

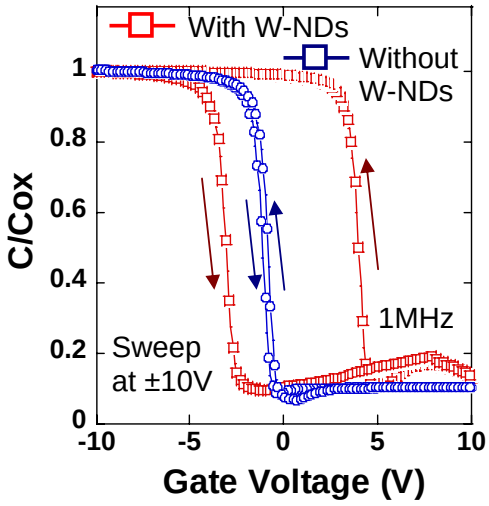


Figure 5. High-frequency (1MHz) C-V curves for MOS memory capacitors with and without W-NDs.

memory capacitor is attributed to the charge exchanging between W-NDs and silicon substrate. We summarized the flat band voltage as a function of sweeping gate bias voltage for W-NDs MOS memory capacitor in Fig. 6. It is found that programming and erasing are from Vg=±4V. At Vg=±5V, the memory window of ~1.0 V was obtained. In general, for the case with silicon oxide control gate, the thickness is designed as 2times as tunneling oxide. From the analysis of C-V curve for capacitor without W-NDs layer, the dielectric constant of HfO_2 was estimated about ~$20\varepsilon_0$ and equivalent oxide thickness (EOT) is about ~8nm. The HfO_2 blocking dielectric with high dielectric constant leads to the coupling to the tunneling oxide effectively. The lower programming voltage and larger memory window were observed. With increasing the gate voltage, the memory window is increased. At middle gate voltage of ±12V, the memory window of 11.86V was obtained. The propriety memory window and power consumption indicate the

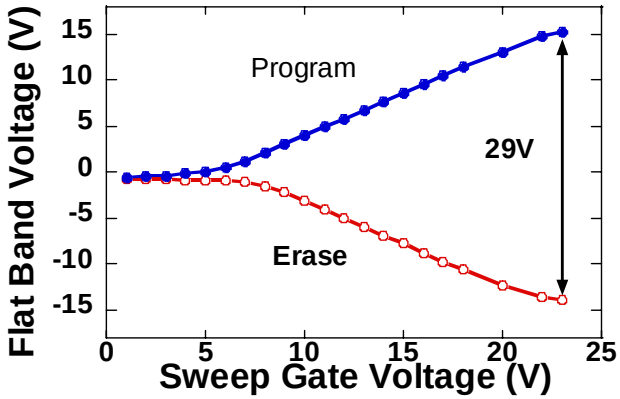


Figure 6. Flat band voltage as a function of sweep gate bias for W-NDs MOS memory capacitor.

suitable candidate for the next generation NVM application. The maximum memory window of 29 V was obtained at ±23V. The large memory window shows the possibility for multi-bit NVM application. And then, the memory window tends to decrease or eliminate due to stress-induced leakage current (SILC) or dielectric breakdown. We have calculated the electron density charged in W-NDs. At memory window of ~29V, the extremely high trapped charge density was estimated about ~$7.8x10^{13}/cm^2$. Considering the dot density and 3D distribution, the result indicates that one W-ND can store 2 or more electrons.

Conclusion

The W-NDs with extremely high density of ~$1.3x10^{13}/cm^2$ and small sizes of 1~1.5nm were dispersed in silicon nitride formed by SAND method. XPS results show that the high metallic states of tungsten were remained after 800℃ anneal due to the surround silicon nitride. Thus high density W-NDs and high-k HfO_2 blocking dielectric were used to investigate the memory characteristics. A large hysteresis memory window of ~11.86V at a middle sweeping gate voltage of ±12V was obtained. The maximum memory window of 29V was obtained at sweeping gate voltage of ±23V, which corresponds to the trapped charge density of ~$7.8x10^{13}/cm^2$, indicating that one W-ND can store 2 or more electrons. These results suggested that memory with floating gate of W-ND in silicon nitride and high-k blocking dielectric becomes a good candidate for future NVM application.

Acknowledgments

This work was partly supported by a Grant-in-Aid. for Scientific Research on Priority Area (No. 18063002) from the Ministry of Education, Culture, Sports, Science, and Technology in Japan.

REFERENCES

1. B. Li, J. Liu, G. F. Liu and J. A. Yarmoff, *Appl. Phys. Lett.*, **91**, 132107 (2007).
2. Yoo-Sung Jang, Jong-Hwan Yoon and Robert G. Elliman, *Appl. Phys. Lett.* , **92**, 253108 (2008).
3. Sung-Soo Yim, Moon-Sang Lee, Ki-Su Kim and Ki-Bum Kim, *Appl. Phys. Lett.*, **89**, 093115 (2006).
4. S. Maikap, T. Y. Wang, P. J. Tzeng, C. H. Lin, L. S. Lee, J. R. Yang and M. J. Tsai, *Appl. Phys. Lett* ., **90**, 253108 (2007).
5. Y. Pei, T. Fukushima, T. Tanaka and M. Koyanagi, *Jpn. J. Appl. Phys.* , **47**, 2680 (2008).
6. Wei-Ren Chen, Ting-Chang Chang, Po-Tsun Liu, Po-Sun Lin, Chun-Hao Tu and Chun-Yen Chang, *Appl. Phys. Lett* ., **90**, 112108 (2007).
7. B. De Salvo et al. , *IEDM*, 999 (2003).
8. Jing Hao Chen, Ying Qian Wang, Won Jong Yoo, Yee-Chia Yeo, Ganesh Samudra, Daniel SH Chan, An Yan Du and Dim-Lee Kwong, *IEEE Transactions on Electron Devices*, **51**, 1840 (2004).
9. Y. Pei, M. Nishijima, T. Fukushima, T. Tanaka and M. Koyanagi, *Appl. Phys. Lett.* , **93**, 113115 (2008).
10. T. H. Fleisch et al., *J. Chem. Phys.*, **76** (1982) 780.

Ex. 1016, Page 65

ECS Transactions, 18 (1) 39-42 (2009)
10.1149/1.3096424 © The Electrochemical Society

Electrical Characterization of W/HfO_2 MOSFETs with La_2O_3 Incorporation

Hiroki Fujisawa[1], A. Srivastava[3-4], Kuniyuki Kakushima[2], Parhat Ahmet[1], Kazuo Tsutsui[2], Nobuyuki Sugii[2], Takeo Hattori[1], C. K. Sarkar[4]and Hiroshi Iwai[1]

[1] FRC, Tokyo Institute of Technology
4259 Nagatsuta, Midori-ku, Yokohama 226-8502, Japan
[2]IGSSE, Tokyo Institute of Technology
[3]Birla Institute of Technology and Science-PILANI
[4]Jadavpur University
fujisawa.h.aa@m.titech.ac.jp

Electrical characterization of HfO_2 MOSFETs with La_2O_3 incorporation has been conducted. A shift in threshold voltage to negative direction has been observed with La_2O_3 incorporation. Moreover, g_m and subthreshold slope have been improved with La_2O_3. 1/f noise measurement has revealed that HfO_2 gated MOSFET incorporating La_2O_3 layer reduces the noise level .

Introduction

Hf-based-oxides have been used to replace SiO_2 gate insulator for advanced CMOS technologies. Recently, La_2O_3, one of rare earth oxides, has attracted much attention as it has wide bandgap of 5.5eV and high dielectric constant of 24 with fairly nice interface properties**.** Also, La_2O_3 has been used to control the threshold voltage to negative direction [1]

In this paper, we investigated the effect of La_2O_3 incorporation underneath the HfO_2 layer on electrical properties of MOSFETs .

Device fabrication and characterization

HfO_2 gated N-channel MOSFET (W/L=50μm/2.5μm) with La_2O_3 incorporation was fabricated on a p-Si (100) with LOCOS isolation. The source and drain were pre-formed prior to the high-k deposition. No interfacial layer was intentionally formed before high-k deposition. 2-nm-thickness of HfO_2 and La_2O_3 with thickness of 4nm were deposited by e-beam evaporation in an ultra high vacuum chamber. Tungsten metal electrode was deposited on top of the high-ks by sputtering. The device structure is shown in fig 1(b). Annealing temperature was performed in forming gas ambient at 500 ^oC for 30 min.

In the same way, a MOSFET with 4-nm-thick-HfO_2 was fabricated as a reference. The equivalent oxide thickness of the MOSFET were 1.6 nm.

The process flow is summarized in fig 1(a) The electrical characteristics of the fabricated high-k gated MOSFETs were measured using Agilent4156C semiconductor parameter analyzer. 1/f noise measurements[2] were carried out with a Agilent HP89410A vector signal analyzer.

Ex. 1016, Page 67

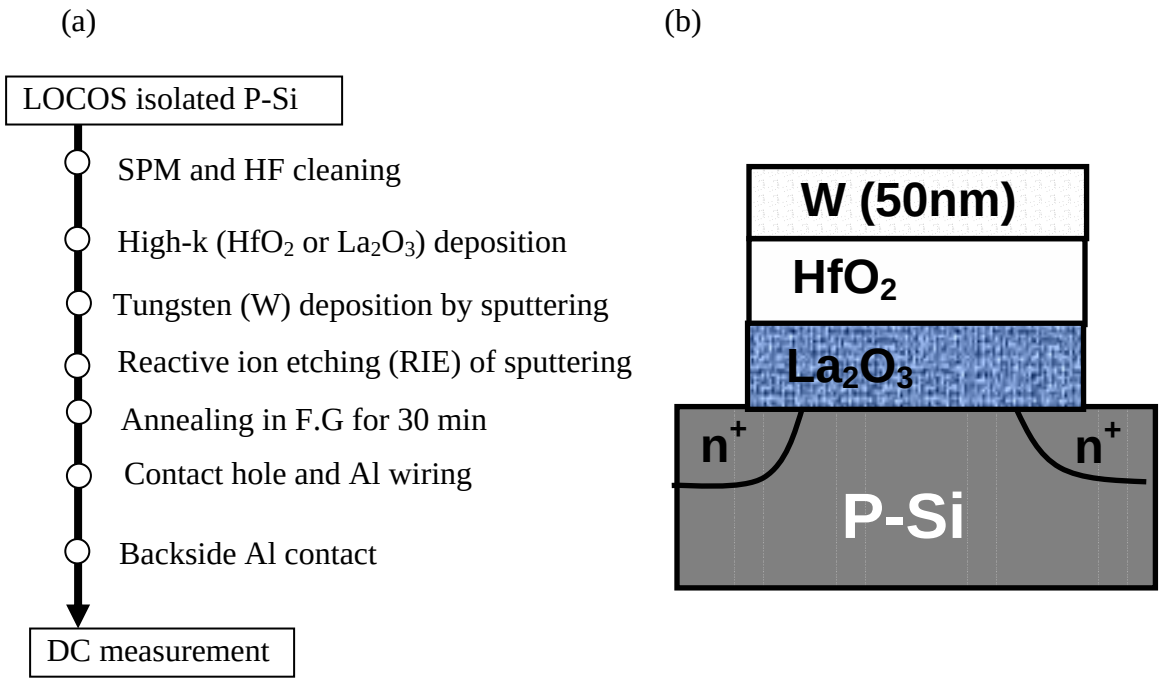
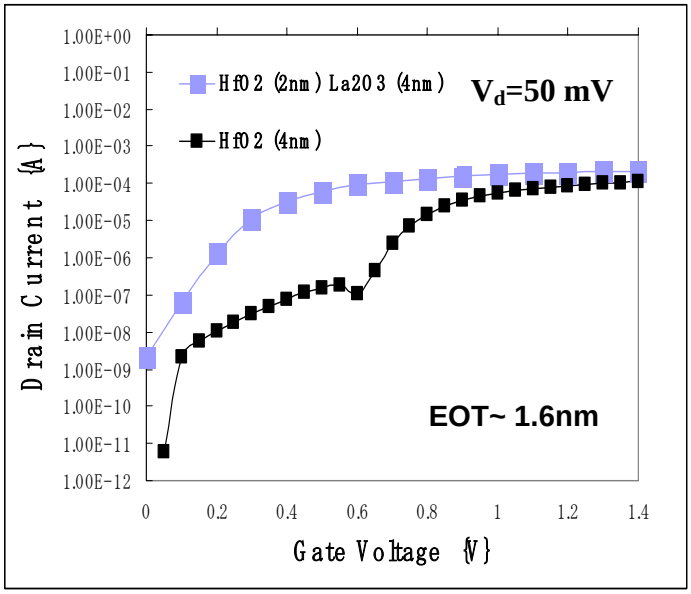


Figure 1. (a) Fabrication process flow of high-k gated n-MOSFET (b) Structure of fabricated MOSFET

Result and Discussion

I_d-V_g of the fabricated MOSFETs are shown in fig 2(a). The threshold voltage was shifted to negative direction by 0.45V with La_2O_3 incorporation.

(a)



(b)

	Vt (V)	SS (mV/dec.)
HfO_2 (4nm)	0.68	94.6
Hf (2nm) La (4nm)	0.23	71.8

Figure 2. (a) I_d-V_g characteristics (b) parameter of these MOSFETs

Fig 2(b) shows the parameter of fabricated MOSFETs. Subthreshold swing (SS) of 94.6 mV/dec was improved to 71.8 mV/dec with the incorporation of La_2O_3 presumably owing to better interfacial property.

The effective electron mobility (μ_{eff}) of the fabricated MOSFET is evaluated by split CV method [3]. Figure 3 shows the calculated μ_{eff} also with HfO_2 and La_2O_3 MOSFETs that have the same EOT of 1.6 nm. The peak mobility was improved from 150 to 200 cm^2/Vs by incorporating La_2O_3 underneath HfO_2 layer.

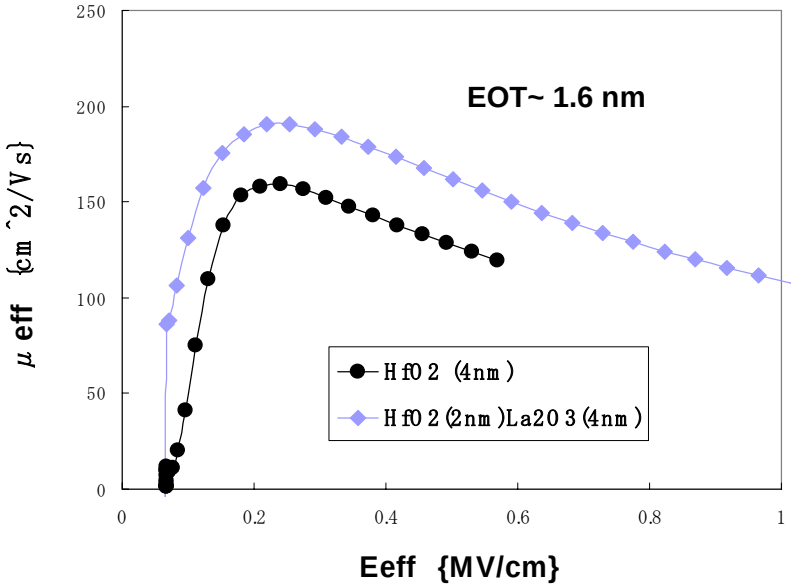


Figure3. Effective electron mobility of La_2O_3 incorporated HfO_2 MOSFET. Samples have the same EOT of 1.6 nm.

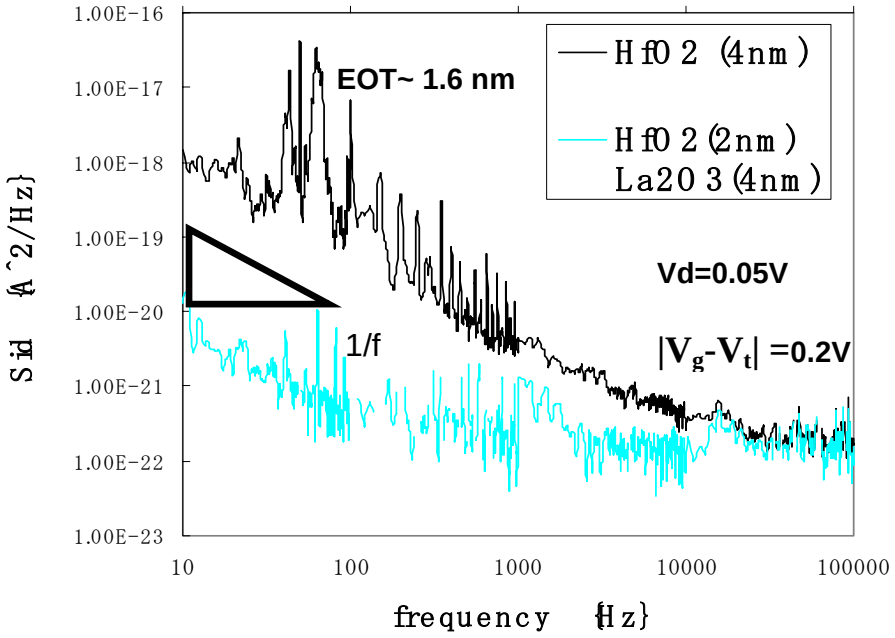


Figure.4 Drain Current Noise vs frequency at V_d=0.05V and $|V_g-V_t|$=0.2V

Figure4. shows the drain current noise (S_{id}) characteristics of the devices at V_d=0.05V and $|V_g-V_t|$=0.2V for n-MOSFETs with W/HfO_2(4nm) /Si and W/La_2O_3(4nm) HfO_2(2nm)/Si In this figure, Sid of the MOSFET incorporated 4nm thickness of La_2O_3 was found to be lower than W/HfO_2(4nm) /Si. Noise level found to be reduced by incorporating La_2O_3 underneath HfO_2 layer.

Conclusions

By incorporating La_2O_3 under HfO_2 layer, a shift in threshold voltage to negative direction was observed. Moreover, the characterization of MOSFET was improved and the peak mobility was improved from 150 to 200 cm^2/Vs. HfO_2 gated MOSFET with incorporating La_2O_3 layer showed lower 1/f noise characteristics.

References

1. K. Tatsumura, *et-al*, *IEEE EDL.*,p25-p28 (2008)
2. P. Srinivasan, ECS Transaction,13(1) 163-169 (2008)
3. C. G. Sodini, T. W. Ekstedt, and J. L. Moll, “Charge Acccumulation and Mobility in Thin Dielectric MOS Transistors”, Solid-State Electron., Vol.25,pp.833-841(1982).

Ex. 1016, Page 70

ECS Transactions, 18 (1) 43-48 (2009)
10.1149/1.3096425 © The Electrochemical Society

The impact of shallow trench isolation on the channel hot electron programming behaviors of SONOS cells

Zhiguo Li, Fan Yang, Joshua Wang, Peter Lin, Jianguang Chang, Yonggang Wang, Champion Yi and Feng Yan
Department of NROM technology development, SMIC, Shanghai 201203, China

Abstract: As scaling down of the flash memory cell size continues, the impact of shallow trench isolation (STI) on active area (AA) in cell array becomes much more severe. Two types of influence have been reported: i) boron segregation: boron atoms at AA corner around STI tend to diffuse out, which induces lower boron concentration at AA corners, generating a low boron concentration region; ii) STI stress: a compressive stress, coming from STI, can extend into AA about 300nm. The former may cause lower threshold voltage of edge cells. The latter will reduce the electron mobility, leading to higher threshold voltage for n-channel cells. In this work, we have studied the impact of STI on the channel hot electron (CHE) programming behaviors of SONOS cells on single cell and array level. It has been found that the programmability of cells adjacent to STI is different from that of center cells. Boron segregation will substantially decrease the program efficiency while the stress will increase the initial programming voltage. In this paper, the impact of STI on edge cells will be reported and compared with the simulation results. The corresponding solutions will be discussed.

INTRODUCTION

As the cell size continues scaling down，Silicon Oxide Nitride Oxide Silicon (SONOS) structures is replacing the traditional floating gate structures, becoming the cell structures of main stream non-volatile memory. Shallow Trench Isolation (STI), as an indispensible element for isolation purpose in deep sub-micron VLSI processing, has been broadly used in non-volatile memory cells in the isolation of array-to-array, contact-to-contact, etc. It has been reported that STI can cause boron loss[1] and introduce compressive stress into Si[2,3], which could impact the performance of non-volatile memories. Though, the impact has been barely studied. In this work, the influence of STI on of SONOS type non-volatile cells has been investigated and analysed based on the simulation and experimental results, especially on the programmability using Channel Hot Electrons(CHE). The corresponding solutions will be presented.

EXPERIMENTS

In this study, NROM architecture has been used, whose cells are formed by the cross stack of bit lines (BL) poly and word lines (WL) poly[4,5]. Figure 1 shows the topview, in which the horizontal lines and the vertical ones are WLs and BLs, respectively. We included 32 WLs and 34 BLs in each array in our experiment. In each array, there are

Ex. 1016, Page 71

two blocks separated by center contact regions where the BL contacts are designed. Between two adjacent BLs, STIs have been added for isolation purpose. The WL width and spacing are 75nm and 125nm. The BL width is 60nm. The gate length is

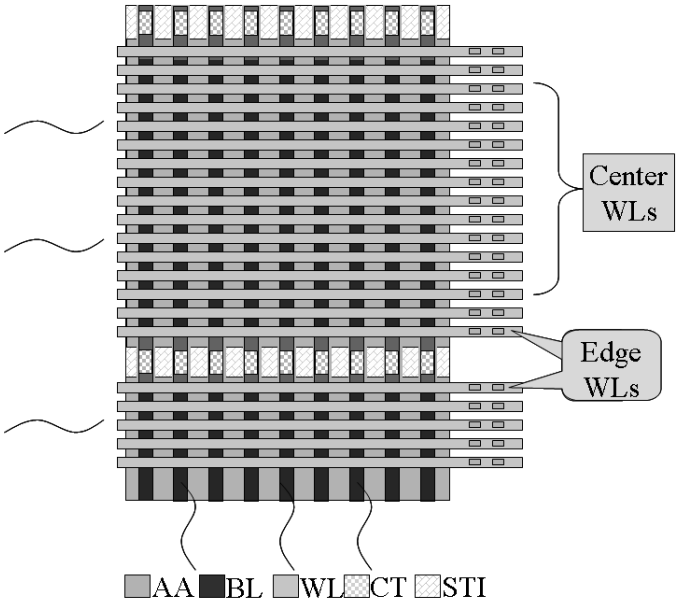


Figure 1 NROM cell array instruction

140nm. Like typical SONOS flash cell, oxide-nitride-oxide (ONO) has been used to replace the gate oxide. The charge is stored in about 5nm nitride layer. In programming, the BL on the drain side should be biased at ~ 5V the BL on the source side is grounded, the gate is forced at 9V, and the substract is grounded. CHE will be generated around the drain, the electrons with energy higher than 3.2eV are injected into nitride layer. As a result, the threshold voltage (V_t) of the drain side increases.

In our study, DC programming test has been conducted to evaluate the effect of the D_{ws}, the distance from the edge WL to STI, on the single cell level. During DC programming test, we force a constant voltage, about 9V, on gate, and then sweep drain voltage, V_d, on the drain from 0V to 7V. The V_d -I_d curve represents the DC programming behavior of single cells. On the array level, V_t distribution has been measured to understand the location dependence and collect statistical V_t data. In addition, Ready-Busy (RB) time, corresponding to the overall programming time on the array level, has been monitored to check overall programming efficiency.

In order to study the impact of STI on the CHE programming behaviors, two kinds of D_{ws}, the distance from edge cells to STI (see Fig. 1), are designed, which are 85nm and 215nm, respectively. We also investigate two types of cells in the array, the edge cells and center cells, i.e. the cells under the WL adjacent to STI and the BL contact region, and the cells under center WLs .

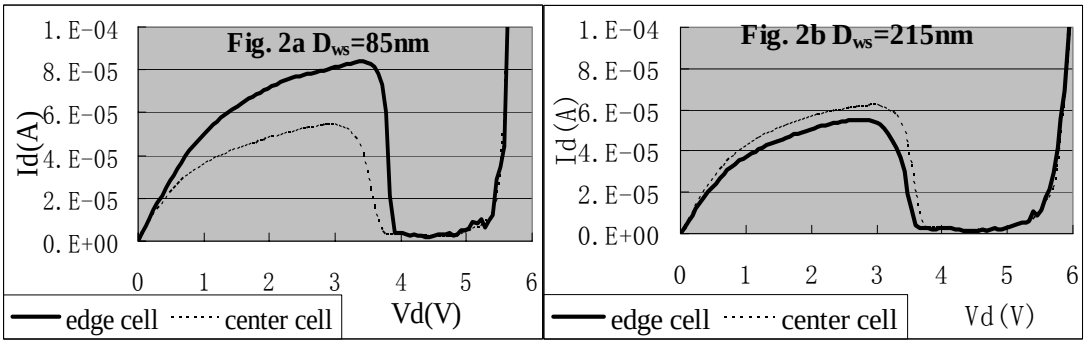
RESULTS AND DISCUSSION

1. DC programming test

From the single cell V_d-I_d curve (Fig. 2a and 2b), it can be seen that the drain current increases as V_d increases. There is a sudden drop on I_d when V_d increases to about 4V. Then I_d will stay at less than $5x10^{-6}$A until 5~6V. The rapid drop of I_d means that the single cell has been programmed. As a result, the V_t of this single cell is much higher than the initial V_t. The channel on the drain side is turned off. The dramatic increase

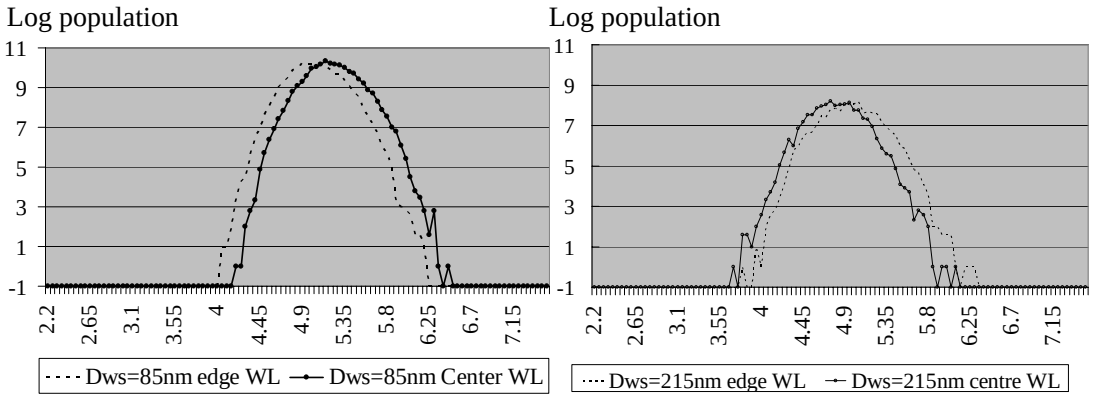
of the I_d occurs above, corresponding to the punch through of the cells.

Fig. 2a tells that edge cells with D_{ws}=85nm need higher programming current and voltage than center cells. But when the extension is increased from 85nm to 215nm, as shown in Fig. 2b, the DC programming curves of center and edge cells merge into one group. No obvious edge and centre cells difference can be observed. Comparing with D_{ws}=85nm, the programming voltage of edge cells is reduced by about 0.2V. The programming current at this point is about 30uA less. As the result, edge cells in the baseline structure with D_{ws}=85nm are harder to be programmed than center cells. In contrast, the difference disappears when D_{ws} is increased to 215nm.



(a) (b)

Figure 2 Single cell DC program curve



(a) (b)

Figure 3 V_t distribution in array level

2. Array level V_t distribution

On the array level, the V_t distribution of center and edge cells with different D_{ws} can be seen in Fig. 3a and 3b. Fig. 3a shows the edge and center cells V_t distribution with D_{ws}=85nm. The V_t of edge cells is 0.2V lower than that of center cells. In contrast, when D_{ws} is extended to 215nm, the edge cells V_t becomes 0.1V higher than that of center cells instead (Fig. 3b).

3. Array level RB time in programming

Based on the RB time results of array level, it has been also found that edge cells show low program efficiency, as shown in Fig. 4. The average programming time of edge cells with D_{ws}=85nm is about 500us more than that of center cells, suggesting that edge cells have lower CHE injection efficiency

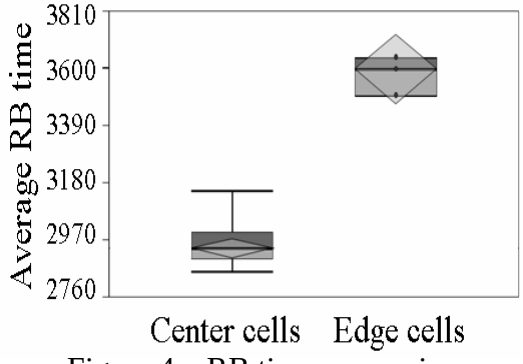


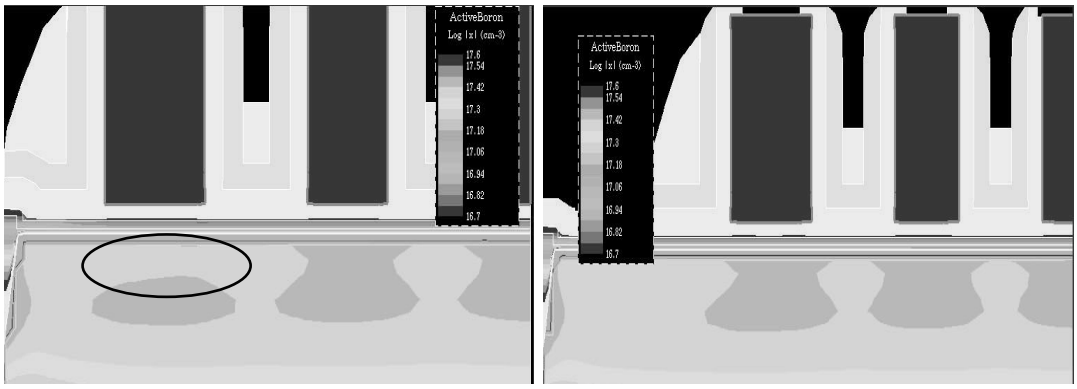
Figure 4 RB time comparison between centre and edge cells

than center cells.

All of the experimental results show that the programming behavior of the edge cells is very different from the center cells. Additionally, the difference changes as D_{ws} changes.

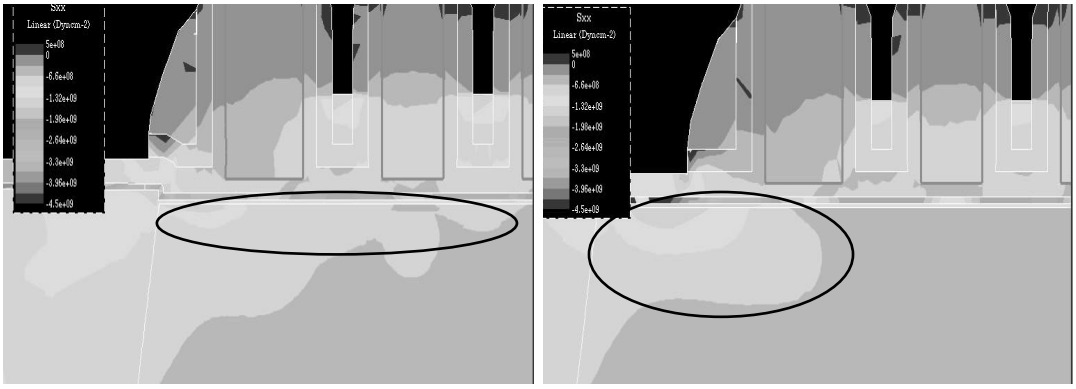
It has been reported that boron atoms at AA corner around STI tend to diffuse out, namely boron segregation[6], inducing lower boron concentration at AA corners. In addition, there exists a compressive stress from STI toward Si. In order to clarify the observed experimental results, Synopsys Tsuprem4 TCAD has been used to simulate the stress and boron distribution. Fig. 5 shows the boron distribution across the cells channel along the BL direction. Fig. 6 shows the horizontal stress distribution along the same direction.

Fig. 5a is the D_{ws}=85nm boron distribution profile and Fig. 5b is the D_{ws}=215nm one. The scale of both profiles ranges from 10^{16} atom/cm^3 to 10^{18} atom/cm^3. In Fig. 5a, the boron concentration in the circle area under edge cells is about 10% lower than that of the adjacent cells, indicating there exists a low boron concentration region in edge cell channel when D_{ws}=85nm due to boron segregation. In contrast, edge cell boron concentration in Fig. 5b is same as center cells when D_{ws} increases to 215nm.



(a) (b)

Figure 5 Boron concentration distribution simulation by Tsuprem4



(a) (b)

Figure 6 Horizontal stress distribution simulation by Tsuprem4

Fig. 6 illustrates the horizontal stress distribution of D_{ws}=85nm (Fig. 6a) and D_{ws}=215nm (Fig. 6b), where the profiles represent compressive stress area. From the stress contour, it can be seen that the channel of the edge cells in the circle has been

impacted by the compressive stress for both D_{ws} =85nm and 215nm. The edge cells suffers about 20% higher compressive stress than center cells (right cells beside circle). The stress can extend into active area (AA) about 300nm, which agrees with the early results[2-3]. It is well known that, the mobility of the electron in the channel decreases under the influence of compressive stress, giving rise to an increase in V_t for n-channel devices.

As seen in Fig. 5 and 6, when D_{ws} is as short as 85nm, the influence on edge cells are from both boron segregation and the stress effect. When D_{ws} is increased to 215nm, boron segregation becomes negligible. Though, the compressive stress from STI to AA dominates the difference of center and edge cells.

Combining the Fig. 5 and Fig. 6 results, the abnormal programming performance of edge cells can be explained as follow: for the edge cells with D_{ws} =85nm, the n-channel boron concentration is substantially lower due to the boron segregation, which results in a lower local electric field at drain side under the same drain bias in turn. Consequently, the program efficiency at edge cells is reduced. Edge cells thus require higher drain biases than center cells in order to be programmed to the same V_t level, which has been clearly shown in the DC programming test Fig. 2 and Fig. 4. When D_{ws} is extended to 215nm, the spacing is long enough to eliminate the impact of boron segregation on the cell channel. Therefore the edge cells performance is the same as the center one. In the same way, RB time can be explained.

The V_t distribution for D_{ws} =215nm shows that the edge cell V_t is slightly higher than center cell one. The reason is the compressive stress from STI to AA. As D_{ws} increases to 215nm, boron segregation can barely affect the edge cells, as shown in Fig. 5. If there did not exist other factors, center cells should have the same Vt as edge cells. Though, the compressive stress from STI to silicon can extend to 300 nm, as shown in Fig. 6. Even when D_{ws} increases to 215nm, the impact of stress still exists. The compressive stress reduces electron mobility, giving rise to the higher V_t on edge cells even as D_{ws} increases to 215nm.

Note that the program efficiency of edge cells is the same for D_{ws} of 85nm and 215nm. This indicates that the program efficiency will not be affected by the channel mobility and the stress from STI. It is only sensitive to the boron concentration in the channel.

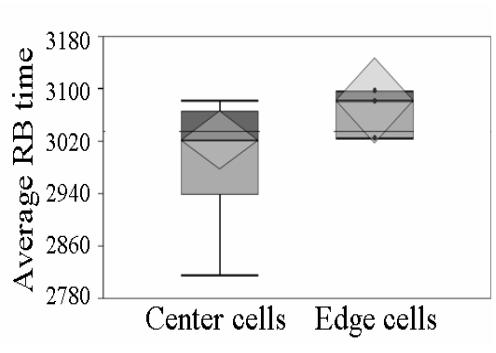


Figure 7 RB time comparison between centre and edge cells

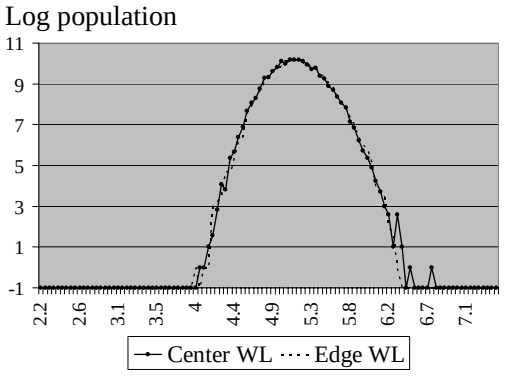


Figure 8 VT distribution in array level

In order to minimize the difference between edge and center cells, additional boron has been implanted in the contact regions to compensate boron segregation around STI as a solution. Fig. 7 and 8 show the RB time and V_t distribution of center and

edge cells with the additional boron implantation. It can be clearly seen that the program efficiency difference has been substantially minimized and the difference of V_t distribution between center and edge cells has been eliminated. The detail results will be reported elsewhere.

CONCLUSION

STI influence on SONOS cells using CHE programming has been investigated. Two factors have been studied: i) boron segregation: boron atoms at AA corner around STI tend to diffuse out, inducing lower boron concentration at AA corners; ii) STI stress: a compressive stress, coming from STI, can extend into AA about 300nm. The former has been identified as the root cause of lower threshold voltage and lower program efficiency of edge cells. The latter will result in higher threshold voltage at edge cells than at center cells. Though, it has little effect on programming efficiency. By imposing an additional implantation in the contact region, the difference between edge and center cells can be minimized.

REFERENCE

1. P.Kuo, J. L. Hoyt, and J. F. Gibbons, “Effects of strain on boron diffusion in Si and SiGe”, Appl. Phys. Lett., vol. 66, No. 5, p. 580, 1995.
2. Masafumi Miyamoto, Hiroyuki Ohta, and Yukihiro Kumagai, “Impact of Reducing STI-Induced Stress on Layout Dependence of MOSFET Characteristics”, IEEE Transactions on Electron Devices, vol. 51, No. 3, p. 440, March 2004
3. Wan Rosmaria, Wan Ahmad, Albert Victor Kordesch, Ibrahim Ahmad, Philip Tan Beow Yew, “TCAD Simulation of STI Stress Effect on Active Length for 130nm Technology”, p. 1040, ICSE2006 Proc., 2006
4. Boaz Eitan, Paolo Pavan, Ilan Bloom, “NROM: A Novel Localized Trapping, 2-Bit Nonvolatile Memory Cell”, IEEE Electron Device Letters, vol. 21, No. 11, p. 543, November 2000
5. Eli Lusky, Yosi Shacham-Diamand, “Investigation of Channel Hot Electron Injection by Localized Charge-Trapping Nonvolatile Memory Devices”, IEEE Transactions on Electron Devices, vol. 51, No. 3, p. 444, March 2004
6. H. Park, K. S. Jones, J. A. Slinkmanm, and M. E. Law, “The effects of strain on dopant diffusion in silicon”, IEDM Tech. Dig., p. 303, 1993.

ECS Transactions, 18 (1) 49-54 (2009)
10.1149/1.3096426 © The Electrochemical Society

Bistable resistance switching of Cu/Ti:ZrO_2/Pt for Nonvolatile memory application

Qi Liu [a, b], Shibing Long [a], Qin Wang [a], Ming Liu [a], Sen Zhang [a], and Junning Chen [b]

[a] Laboratory of Nana-Fabrication and Novel Devices Integrated Technology, Institute of Microelectronics, Chinese Academy of Sciences, Beijing 10029, China
[b] College of Electronics and Technology, Anhui University, Hefei 230039, China

Resistive switching characteristics of Ti doped ZrO_2 films are investigated for nonvolatile memory applications. The devices are with a metal-insulator-metal structure adopting Ti doped ZrO_2 films sandwiched between top Cu electrode and bottom Pt electrode. Both DC and pulse voltage can trigger the switching between high resistance state and low resistance state. The ratio between the high and low resistance can be on the order of 10^4 after more than 250 switching cycles. In voltage pulse mode, "write" and "erase" operation can be as fast as 5 ns and 100 ns, respectively. Furthermore, no data loss is found for more than 10^6 seconds. Finally, the switching mechanism is exploited and concluded to be related with the formation and rupture of conducting filamentary paths.

I. INTRODUCTION

Recently, as one of the candidates of the next-generation nonvolatile memories, resistance random access memory (ReRAM) has attracted extensive attention due to its superior characteristics including simple structure, excellent scalability, high density, low power, and fast speed (1). Large variety of materials such as ferromagnetic materials (2), organic materials (3), doped perovskite (4) and binary transition metal oxides (5-9, 12-13) are proved to possess the resistance switching characteristics under certain external stimulus. Among these, binary transition metal oxides are preferred due to their simple structure, easy fabrication process, and compatibility with standard CMOS process (1). The defects and impurities in the metal oxide which play a critical role in resistive switching have been recognized, though some controversies regarding resistive switching mechanisms still exist. Recently, the intentionally introduced external impurities in metal oxide to improve switching performance have attracted wide attention (5, 8, 9). In our previous study, we demonstrated that the implanted Zr^+ in zirconium oxide can greatly improve the on/off ratio and device yield (5). However, the endurance characteristic of Zr^+-implanted-ZrO_2 film still needs to be optimized.

In this study, we adopt Ti as external impurity in order to improve switching properties of the ZrO_2-based resistive switching memory. And it was introduced by ion implantation technique. Both bipolar and unipolar resistance switching could be observed in the device. The Cu/Ti:ZrO_2/Pt device showed high device yield, large resistive ratio, fast speed, good endurance and long retention characteristics. The physical origin of the switching phenomena was also suggested.

Ex. 1016, Page 77

II. EXPERIMENT

The resistive switching memory devices in this study were fabricated as follows. After thermal oxidation of silicon wafer (1000 °C, 5 hours, oxygen flow rate: 6.7 L/min), 20 nm Ti and 80 nm Pt bottom electrode were deposited on the SiO_2/Si substrate via e-beam evaporation. The Ti film was used for adhesion layer. Then, 70 nm ZrO_2 film was deposited through electron beam evaporation with 1 Å/s deposition rate, under pressure of 2.6×10^{-6} Torr. Followed annealing was carried out at 800 °C for 120 s in N_2 ambient (2.5 L/min). Thereafter, Ti ions were implanted into the ZrO_2 film with 1×10^{12} cm^{-2} dose and 50 KeV energy. To activate Ti ions and repair lattice damage, Ti doped ZrO_2 film was annealed at 400 °C for 120 s in N_2 ambient (2.5 L/min). Finally, Cu (70 nm) and Au (30 nm) layers as top electrode were deposited subsequently. The Au layer was to avoid Cu electrode oxidation during testing and prevent device surfaces from being scratched by the probe tip. All top electrodes were defined as squares (area from 50×50 μm^2 to 800×800 μm^2) after lift-off process. The DC current-voltage (I-V) characteristics of the fabricated devices were measured using the Keithley 4200 semiconductor characterization system. The pulse test was performed with the Agilent 81110A Pulse Pattern Generator and the Tektronix DPO7104 Oscilloscope. During the electrical measurements, the bias polarity was defined with reference to the bottom Pt electrode, as shown in inset of Fig. 1 (a).

III. RESULTS AND DISCUSSION

The Cu/Ti:ZrO_2/Pt device showed both bipolar and unipolar resistive switching, as shown in Fig. 1(a) and 1(b). For the pristine device, the initial state was in the off-state and the resistance was in the order of ~10^7 Ω. As can be seen from Fig. 1 (a), a conspicuous bipolar I-V switching was observed in the Cu/Ti:ZrO_2/Pt device under voltage sweeping ((0 V→1 V→0 V→−1 V→0 V). When positive voltage was applied on the top electrode, the device current was abruptly increased at about V_{set} and the device switched from the off-state to the on-state, which was denoted as “Set process”. When sweeping from 0 to -1.0 V, the device current was abruptly decreased at V_{reset} and the device switched from on-state to the off-state, which was denoted as “Reset process”. Note that, the Reset process could also be achieved through applied positive voltage, as shown in Fig 1 (b). In this case, the Set voltage and Reset voltage have the same polarity, which was known as “unipolar” switching. In the following test, Set and Reset process were operated by positive voltage and negative voltage, respectively. This could avoid competition between Set and Reset process (6) due to the Set process was difficult to occur under negative voltage. The resistance of LRS and HRS were ~ 100 Ω and ~10^7 Ω under 0.1 V read voltage, respectively. The resistance ratio between these two states could reach more than 10^5. The resistance switching could be traced for hundreds of times, and more than 90% devices under test showed the same resistive switching behavior. Additionally, resistive switching could be achieved under voltage pulse mode. Set and Reset processes could be as fast as 5 ns and 100 ns, respectively.

Ex. 1016, Page 78

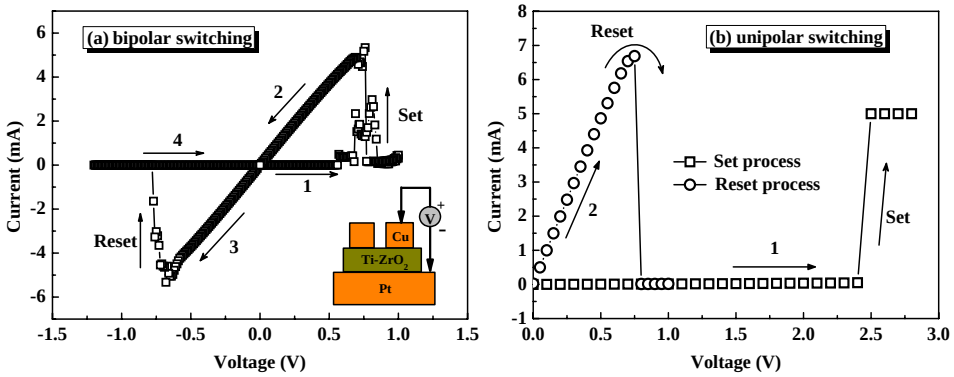


Fig. 1 The typical I-V characteristics of the Cu/Ti:ZrO_2/Pt device. (a) bipolar switching mode and (b) unipolar switching mode. The arrows indicate the sweeping directions, and the schematic diagram of measurement structure was showed in the inset of fig. 1 (a).

To evaluate the potential application of the fabricated Cu/Ti:ZrO_2/Pt devices in nonvolatile memory, we further tested the endurance and retention properties. These properties are the most crucial factors in developing nonvolatile memory devices. Fig. 2 showed the dependence of the resistance on the repetitive Set-Read-Reset-Read cycles in a 100 × 100 um cell. The amplitude of the Set/Reset voltage and readout voltage were ± 6 V and +0.1 V, respectively. In Set process, the compliance current was kept at 1 mA to prevent permanent breakdown of devices. As could be seen from Fig. 2, there was a narrow distribution of resistance values in both LRS and HRS. After more than 250 Set-Read-Reset-Read cycles, the HRS/LRS ratio could still remain at more than 10^4.

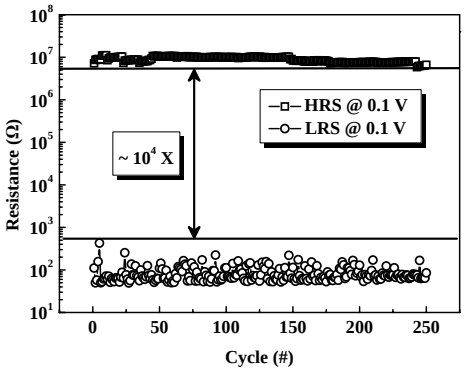


Fig. 2 Endurance performances of the Cu/Ti:ZrO_2/Pt device. The endurance test was performed under a series of 6 V/0.1 V/-6 V/0.1 V voltage cycles. In the Set process, the compliance current was 1 mA.

Fig. 3 showed the retention performance of both LRS and HRS for the Cu/Ti:ZrO_2/Pt device under 0.1 V readout voltage at room temperature. The resistance in both LRS and HRS were mostly unchanged for more than 10^6 seconds under room temperature. After these retention tests, the device still showed reproducible switching behavior without obvious degradation. Fast speed, good endurance, and excellent retention characteristics

of the Cu/Ti:ZrO_2/Pt device implied its promising potential for nonvolatile memory application.

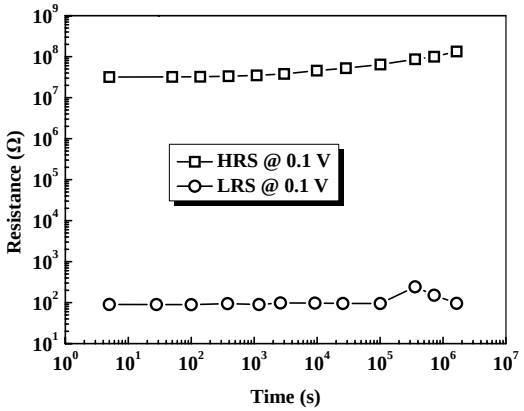


Fig 3. Time dependence of both resistance states at room temperature. The resistance of both LRS and HRS were stable, and there was no obvious degradation over a time span of 10^6 s.

In general, I-V curves have been analyzed to study the underlying mechanism of the resistive switching in memory devices (7). As shown in Fig. 1 (a), it could be seen that the I-V characteristic in the on-state followed a linear ohmic conduction (slope was close to 1). This was probable due to the formation of conducting filaments throughout the insulating film. In order to confirm our assumption, the relationship between resistance and the device area was studied. We obtained each averaged data point by repeating the DC switching cycles over 40 times on different cells. As shown in the Fig 4, the resistance of on-state was insensitive to the cell sizes, indicating that the "on" current flows through localized conducting filamentary paths in the insulator, which confirmed the filament transmission mechanism in the on-state. Additionally, the resistance of the off-state increased with a reduced cell area, which suggested that the Cu/Ti:ZrO_2/Pt device has a larger sensing margin when the device is scaled down.

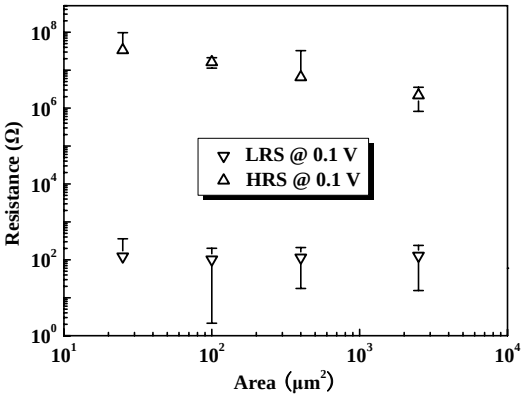


Fig 4. Resistance values of Cu/Ti:ZrO_2/Pt device depended on junction area for both the on-state and off-state.

Recent studies have shown that the conducting filamentary paths in the transition metal oxides could be formed by the percolation of some kind of defects, such as the

migration of metal ions (9) or oxygen vacancies (10). Yan Cong et al. studied the luminescent properties of Ti doped ZrO_2 (11) and found that doping Ti into ZrO_2 may induce more oxygen vacancy due to that Zr^{4+} is replaced by Ti^{3+} in the ZrO_2 crystal lattice. We suggested that the resistive switching phenomena in our device were caused by these additional oxygen vacancies which were induced by the doped Ti impurities. In the set process, the charged oxygen vacancies in ZrO_2 layer migrate towards the anode under electric field, and an oxygen-deficient region starts to build and to extend towards the anode (12). The conductance of these region is changed from oxide into metallically conducting phase due to an abundance of transition metal cations (such as Zr or Ti ions) are generated. This metallically conducting phase region moves towards the anode and finally forms a conductive path. In the reset process, a few conducting filamentary paths mainly conduct the large current. Heating effect around these conducting filamentary paths during the process of measurement is unavoidable, which causes the rise of the local temperature around these filaments. And these filaments will break down in the maximum temperature region (13).

IV. CONCLUSION

In summary, the Cu/Ti:ZrO2/Pt sandwich structure was fabricated and investigated for nonvolatile memory applications. It was found that the Cu/Ti:ZrO_2/Pt device exhibited stable and reproducible resistive switching behaviors. Considering that the Cu/Ti:ZrO_2/Pt device showed high device yield, good endurance, fast speed, and long retention characteristics, it had great potential for application in future nonvolatile resistive switching memory devices. Based on the experiment, the formation and rupture of conducting filamentary paths related to the implanted Ti ions were suggested to be responsible for the resistive switching phenomena.

Acknowledgments

This work was supported by the Hi-Tech Research and Development Program of China (863 Program) under Grant Nos 2008AA031403 and 2008AA031401, the National Basic Research Program of China (973 Program) under Grant Nos 2006CB302706, 2006CB806204 and 2007CB935302, the National Natural Science Foundation of China under Grant Nos 60825403, 90607022 and 60506005 and Chinese Academy of Sciences under Grant No YZ200840.

References

1. I. G. Baek, M. S. Lee, S. Seo, M. J. Lee, D. H. Seo, D. S. Suh, J. C. Park, S. O. Park, H. S. Kim, I. K. Yoo, U-I. Chung, and I. T. Moon, in *IEDM Tech. Dig/2004*, p. 587
2. S. Q. Liu, N. J. Wu, and A. Ignatiev, *Appl. Phys. Lett.*, 76, 2749 (2000).
3. L. Ma, S. Pyo, J. Ouyang, Q. Xu, and Y. Yang, *Appl. Phys. Lett.*, 82, 1419 (2003)..
4. C. C. Lin, B. C. Tu, C. C. Lin, C. H. Lin, and T. Y. Tseng, *IEEE Electron Device Lett.*, 27, 725 (2006).
5. Q. Liu, W. Guan, S. Long, R. Jia, M. Liu, and J. Chen, *Appl. Phys. Lett.*, 92, 012117 (2008).
6. M. Yin, P. Zhou, H. B. Lv, J. Xu, Y. L. Song, X. F. Fu, T. A. Tang, B. A. Chen, and Y. Y. Lin, *IEEE Electron Device Lett.*, 29, 681 (2008).

7. Y. Xia, W. H, L. Chen, X. Meng, and Z. Liu, *Appl. Phys. Lett*., 90, 022907 (2007)
8. K. Tsunoda, K. Kinoshita, H. Noshiro, Y. Yamazaki, T. Iizuka, Y. Ito, A. Takahashi, A. Okana, Y. Sato, T. Fukano, M. Aoki, and Y. Sugiyama, in *IEDM Tech. Dig/2007,* p. 767.
9. W. Guan, M. Liu, S. Long, Q. Liu, and W. Wang, *Appl. Phys. Lett*. 93, 223506 (2008).
10. M. Janousch, G. Ingmar Meijer, U. Staub, B. Delley, S. F. Karg, and B. P. Andreasson, *Adv. Mater.*, 19, 2232 (2007)
11. Y. Cong, B. Lin, B. Lei, W. Li, *J. Luminescence,* 126, 822 (2007).
12. R. Waser, and M. Aono, *Nature Mater.*, 6, 833 (2007)
13. Y. Sato, K. Kinoshita, M. Aoki, and Y. Sugiyama, *Appl. Phys. Lett.*, 90, 033503 (2007)

ECS Transactions, 18 (1) 55-60 (2009)
10.1149/1.3096427 © The Electrochemical Society

Transport in Silicon Nanowire Transistors

T. Hiramoto, J. Chen, Y.J. Jeong, and T. Saraya

Institute of Industrial Science, University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

Mobility and strain effects in silicon nanowire MOSFETs are extensively studied by experiments. The electron mobility of silicon nanowire with width less than 10nm has been evaluated by the split CV method for the first time. It is found that the mobility decreases with decreasing nanowire width and depends strongly on nanowire direction. The strain effects of nanowire nMOSFETs have been evaluated for the first time. It is found that the transverse tensile strain offers more favorable effects than the longitudinal one in terms of Ion/Ioff ratio.

Introduction

The silicon nanowire MOSFET is a promising structure for future VLSIs because of its high short channel effect immunity. When the width of the nanowire is scaled down to the nanometer regime, the transport properties are strongly affected by the quantum confinement effects. The transport is also affected by applying strain. In order to attain high performance nanoscale MOSFETs, the basic understandings of mobility, transport, strain effect, and band structures of silicon nanostructures are indispensable. A new design methodology of nanoscale devices, in which the new physics in nanostructures are positively utilized while disadvantages due to nano physics are minimized, is strongly required (1).

In nanowire MOSFETs (2-8), only a few experimental works have been reported previously on the measurements on mobility and strain effects. The experimental study on quantum confinement effects in nanowire MOSFETs have not been sufficient. In this study, the exact mobility measurements (9) and strain effect measurements (10) are performed systematically and the physical origins of the measured electrical properties are discussed.

Mobility in Silicon Nanowire MOSFETs

The device fabrication starts with thinning a (100) silicon-on-insulator (SOI) wafer to the desired thickness. After nanowire patterns are defined by the electron beam (EB) lithography and followed by reactive ion etching (RIE), the nanowire width is further narrowed using SC1 solution. Then, BHF is used to over-etch the buried oxide (BOX) layer under the nanowires for gate-all-around (GAA) structure.

The final nanowire width (W_{NW}) after gate oxidation is from 7nm~48 nm, with gate oxidation thickness around 13.7 nm on (100)-surface while 21.6nm on (110)-surface. The SOI thickness (T_{SOI}, NW height) and nanowire width are estimated by ellipsometry and scanning electron microscope (SEM) observation. [110]-directed and [100]-directed FETs are fabricated. Ultra-thin-body (UTB) MOSFETs are also fabricated for reference.

Ex. 1016, Page 83

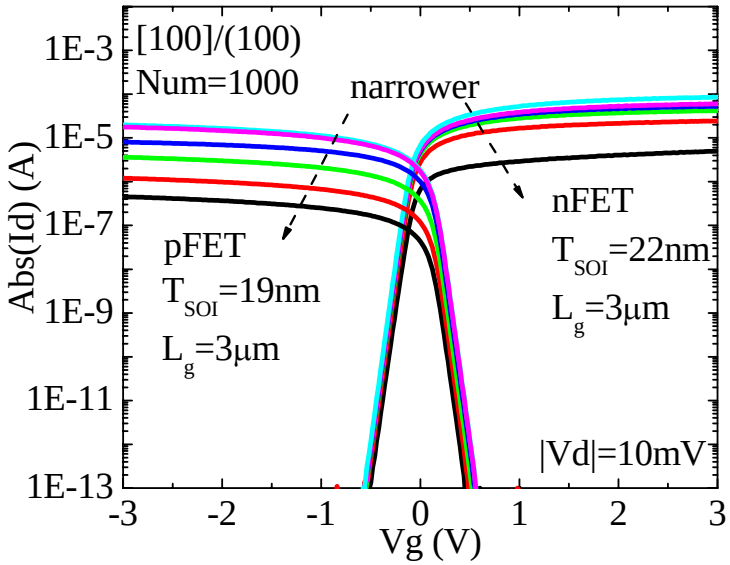


Figure 1. (a) Measured Id-Vg characteristics of [100]-directed nFET and pFET with 1000 nanowires. W_{NW}= 7nm~48nm.

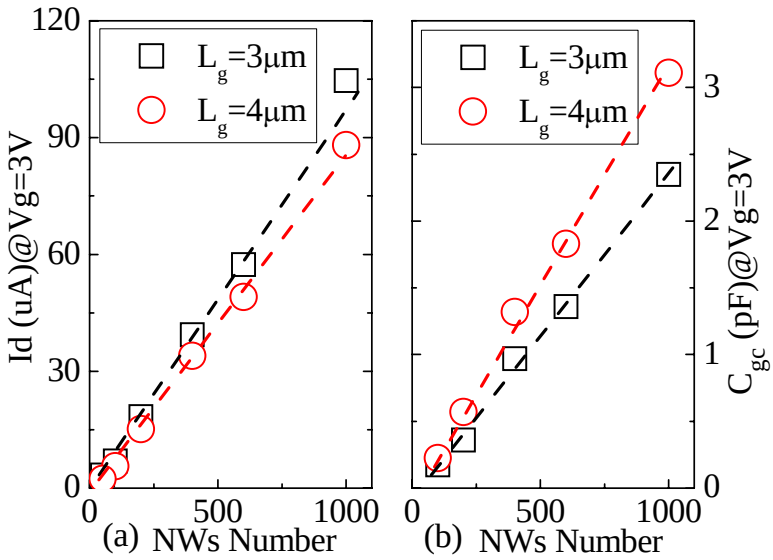


Figure 2. Linear relationships with nanowire number of (a) I-V and (b) C-V characteristics in nanowire nFETs.

Fig. 1(a) shows Id-Vg characteristics of [100]/(100) nanowire MOSFETs with 1000 parallel nanowires (T_{SOI}=22nm and Lg=3μm). The drain current decreases as the nanowires turn to be narrower. Thanks to the GAA structure and surface roughness control, sub-threshold characteristics show ideal performance in the whole nanowire width range (S ~ 61 mV/dec). Fig. 2(a) and (b) show Id and Cgc of 48nm nanowires at Vg=3 V, changing the number of parallel nanowires and gate length (Lg=3μm, 4μm). The linear relationships along with nanowire number are observed clearly. Accurate mobility estimations are performed by using the double L_m method (11), where two gate lengths are designed to remove the effect of parasitic capacitance and resistance. The average surface carrier density is calculated by $N_{inv}= N_l/(2W_{NW}+2T_{SOI})$, where N_l means the line inversion carrier density obtained from C-V measurement.

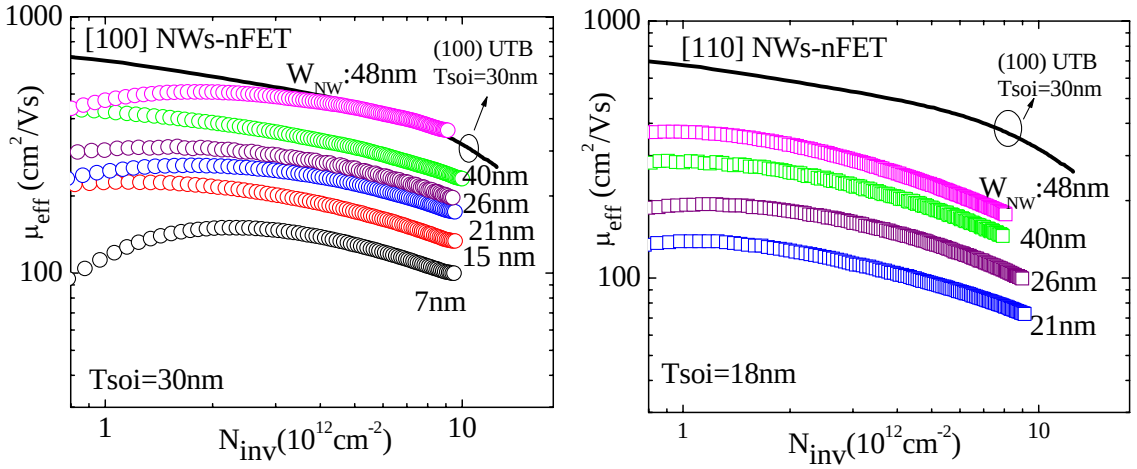


Figure 3. Measured electron mobility of (a) [100]-directed nanowire nFETs with T_{SOI}=30nm and (b) [110]-directed nanowire nFETs with T_{SOI}=18nm.

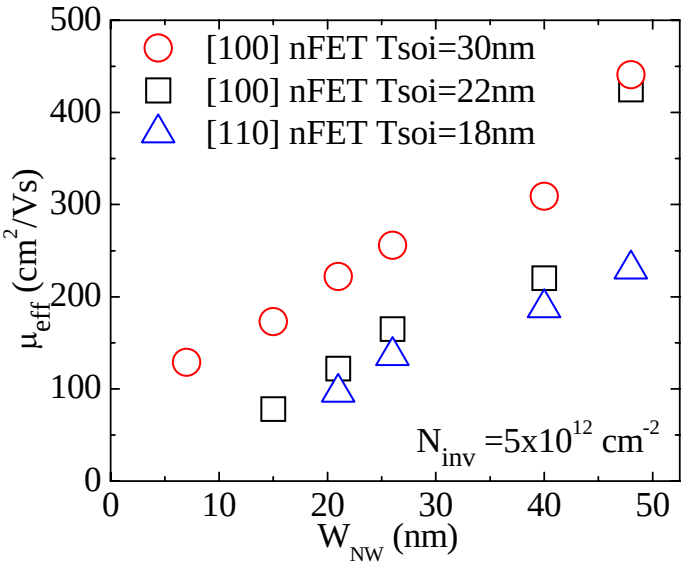


Figure 4. Measured electron mobility at N_{inv}=5x10^{12} cm^{-2} as a function of W_{NW}.

Mobility in [100]-directed nanowire nFETs with T_{SOI} =30nm are shown in Fig. 3(a). In the wider-limit, mobility is even higher than that of UTB at high N_{inv}. The mobility degradation tendency with decreasing W_{NW} is clear. Devices with thinner T_{SOI} have smaller mobility. Mobility in [110]-directed nanowire nFETs with T_{SOI} =18nm is shown in Fig. 3(b). Because electron mobility on (110) surface is lower than that of (100) surface, [110] nanowire has smaller mobility than [100] nanowire with W_{NW}=48nm. However, this degraded mobility in [110] nanowire diminishes and is comparable to that of [100] nanowire as decreasing W_{NW}, as shown in Fig. 4.

As reported in UTB nFET, due to the subband modulations, a (100) double-gate (DG) nFET has smaller mobility than a (100) single-gate (SG), while (110) DG nFET has larger mobility than (110) SG (12). Please note that the [110]-directed nanowire nFET is in form of DG nFET on (110) surface, and [100] nanowire nFET is in form of DG on (100) surface. Therefore, comparable mobility observed in [110] and [100] nanowire nFETs can be explained by the subband modulation effect on (110) surface.

Strain Effects in Silicon Nanowire MOSFETs

Silicon nanowire nFETs with a single nanowire are fabricated on (100) SOI substrate. The nanowire width is varied from ~1 nm to ~20 nm. The standard deviation of the channel width σ is less than 2nm (2). V_{th} (defined as Vg at $I_{d0} = 10^{-12}$ A) increases as nanowire width decreases due to the quantum confinement effect. Uniaxial tensile stress was applied mechanically using a 4-point bending apparatus and estimated with an analytical model. All of the measurements were carried out at room temperature. Drain voltage is fixed to 10 mV.

Fig. 5 shows transverse tensile strain (=90 MPa) effects on the drain current of nanowire nFETs. The vertical axis is the current modulation factor ($\Delta I_d/I_d$) and the horizontal axis is the overdrive voltage ($V_{over} = V_g - V_{th}$). At low V_{over}, negative $\Delta I_d/I_d$ is observed. This is attributed to strain-induced positive V_{th} shift (ΔV_{th}), which is not consistent with bulk MOSFETs. However, as V_{over} gets larger and inversion layer starts to form, current reduction fades and finally $\Delta I_d/I_d$ turns to positive as the strain induced mobility enhancement overcomes the effect of V_{th} increase. At V_{over} higher than ~0.4 V, current improvement resulting from mobility enhancement converges to ~0.7%, even at the narrowest nanowire FET. This tendency is consistent to the bulk case. These results suggest that transverse tensile strain is favorable because it leads to I_{on} improvement and I_{off} reduction.

Fig. 6 shows longitudinal tensile strain (=110 MPa) effects on drain current of nanowire FETs. In contrast to transverse strain, $\Delta I_d/I_d$ begins with positive value at low V_{over}, attributed to V_{th} decrease under longitudinal tensile strain. Up to high V_{over}, current increase is maintained and $\Delta I_d/I_d$ converges to ~0.5%, which means that longitudinal tensile strain effect for mobility enhancement is valid from bulk to ultra-narrow nanowire FETs.

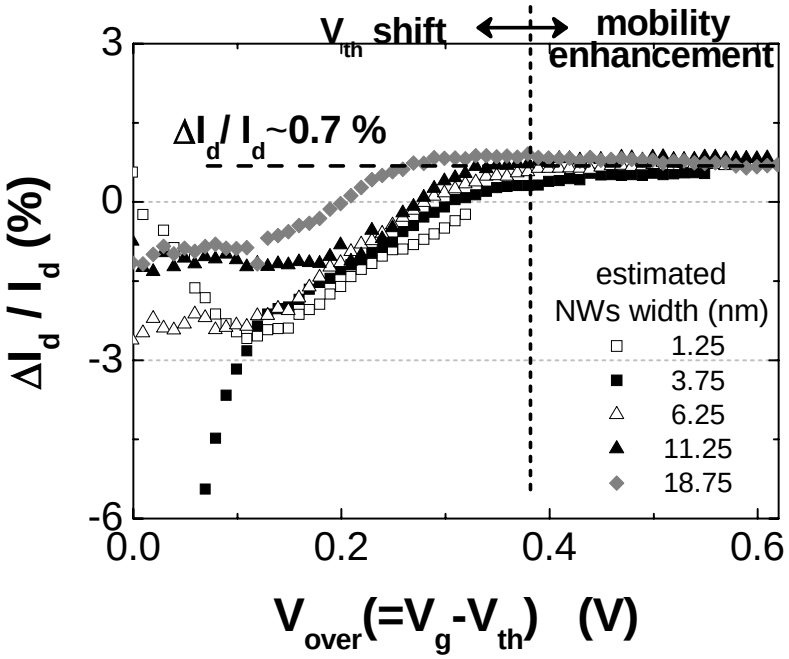


Figure 5. Measured $\Delta I_d/I_d$ - V_{over} characteristics of nanowire nFETs under transverse tensile strain.

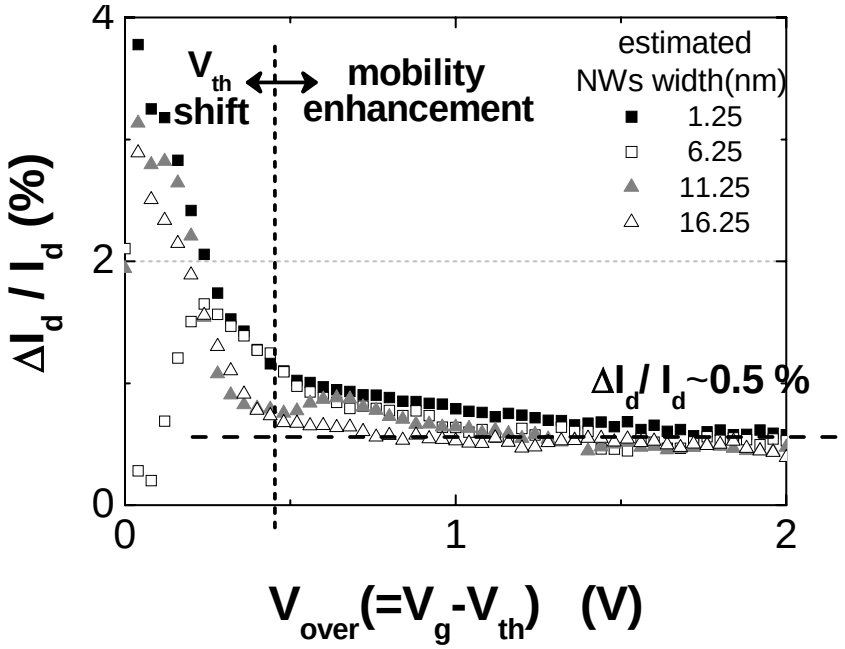


Figure 6. Measured $\Delta I_d/I_d$ - V_{over} characteristics of nanowire nFETs under longitudinal tensile strain.

Conclusion

Mobility in silicon nanowire nFETs with nanowire width less than 10nm and strain effects in nanowire nFETs are systematically measured for the first time. The electrical behaviors of silicon nanowire FETs are modified by quantum confinement effects and basic understanding of these properties are essential for the design of nanoscale MOSFETs in the future. More experimental results on nanowire FETs have recently been presented (13, 14).

Acknowledgments

This work was partly supported by Innovation Research Project on Nanoelectronics Materials and Structures of METI, Japan.

References

1. T. Hiramoto, M. Saitoh, and G. Tsutsui, *IBM Journal of Research and Development*, **50**, 411 (2006).
2. H. Majima, H. Ishikuro, and T. Hiramoto, *IEEE Electron Device Letters*, **21**, 396 (2000).
3. H. Majima, Y. Saito, and T. Hiramoto, IEEE International Electron Devices Meeting (IEDM), p. 733 (2001).
4. F.-L. Yang, D.-H. Lee, H.-Y. Chen, C.-Y. Chang, S.-D. Liu, C.-C. Huang, T.-X. Chung, H.-W. Chen, C.-C. Huang, Y.-H. Liu, C.-C. Wu, C.-C. Chen, S.-C. Chen, Y.-T. Chen, Y.-H. Chen, C.-J. Chen, B.-W. Chan, P.-F. Hsu, J.-H. Shieh, H.-J. Tao, Y.-C. Yeo, Y. Li, J.-W. Lee, P. Chen, M.-S. Liang and C. Hu, Symposium on VLSI Technology, p. 196 (2004).

Ex. 1016, Page 87

5. S. D. Suk, S.-Y. Lee, S.-M. Kim, E.-J. Yoon, M.-S. Kim, M. Li, C. W. Oh, K. H. Yeo, S. H. Kim, D.-S. Shin, K.-H. Lee, H. S. Park, J. N. Han, C. J. Park, J.-B. Park, D.-W. Kim, D. Park, and B.-I. Ryu, IEEE International Electron Devices Meeting (IEDM), p. 735 (2005).
6. J. Wang, A. Rahman, G. Klimeck, and M. Lundtrom, IEEE International Electron Devices Meeting (IEDM), p. 537 (2005).
7. H. Lee, L.-E. Yu, S.-W. Ryu, J.-W. Han, K. Jeon, D.-Y. Jang, K.-H. Kim, J. Lee, J.-H. Kim, S. Jeon, G. Lee, J. Oh, Y. Park, W. Bae, H. Lee, J. Yang, J. Yoo, S. Kim and Y.-K. Choi, Symposium on VLSI Technology, p. 70 (2006).
8. T. Tezuka, E. Toyoda, S. Nakaharai, T. Irisawa, N. Hirashita, Y. Moriyama, N. Sugiyama, N. Taoka, Y. Yamashita, O. Kiso, M. Harada, T. Yamamoto, and S. Takagi, IEEE International Electron Devices Meeting (IEDM), p. 887 (2007).
9. J. Chen, T. Saraya, K. Miyaji, K. Shimizu, and T. Hiramoto, Symposium on VLSI Technology, p. 32 (2008).
10. Y.J. Jeong, K. Miyaji, and T. Hiramoto, IEEE Silicon Nanoelectronics Workshop, 2008.
11. A. Toriumi, K. Kita, and H. Irie, IEEE International Electron Devices Meeting (IEDM), p. 671 (2006).
12. G. Tsutsui, M. Saitoh, T. Saraya, T. Nagumo, and T. Hiramoto, IEEE International Electron Devices Meeting (IEDM), p. 747 (2005).
13. J. Chen, T. Saraya, and T. Hiramoto, IEEE International Electron Devices Meeting (IEDM), p. 757 (2008).
14. Y.J. Jeong, J. Chen, T. Saraya, and T. Hiramoto, IEEE International Electron Devices Meeting (IEDM), p. 761 (2008).

ECS Transactions, 18 (1) 61-66 (2009)
10.1149/1.3096428 © The Electrochemical Society

pMOSFET Off-state Leakage and Junction Leakage Current in Ge-based Devices

E. Simoen[a], F. De Stefano[a,b], G. Eneman[a,c,d], B. De Jaeger[a], C. Claeys[a,c] and F. Crupi[b]

[a] IMEC, Kapeldreef 75, B-3001 Leuven, Belgium
[b] Universita degli Studi della Calabria, Via P. Bucci Arcavacata di Rende (CS), 87100 Italy
[c] EE Depart KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
[d] Post-doc Researcher Fonds Wetenschappelijk Onderzoek Vlaanderen

The impact of the post-implantation Rapid Thermal Anneal (RTA) temperature on the leakage current of p^+n junctions fabricated in epitaxial Ge-on-Si substrates is analyzed in terms of the area and the peripheral leakage current component. While the former is not clearly affected, it is shown that the peripheral leakage current density increases in the range 500 to 600°C. A similar trend is found for the off-state leakage current of pMOSFETs made in the same substrates. This is ascribed to a higher activation of the halo implantation, which generates a higher band-to-band tunneling at the interface. It is, finally, shown that replacing the standard BF_2 extension implantation by a Ge pre-amorphization followed by a low-energy B implant improves the area leakage current density, most likely because of a lower residual damage.

Introduction

One of the concerns in the development of germanium-channel CMOS is the relatively low band gap energy, facilitating the generation of leakage current by defect levels or direct valence-to-conduction-band excitation [1,2]. According to Fig. 1, we can distinguish different components in the drain-to-bulk leakage current, which are associated with the peripheral isolation region (corresponding density J_P in A/cm), the Highly-Doped Drain (HDD) leakage to the bulk (density J_A in A/cm^2) and the leakage associated with the extension regions with density J_E (in A/cm). While it is not evident to study J_E [3], the other components (J_A and J_P) can be fairly easily evaluated by measuring the reverse current (I) versus voltage (V) on p-n junctions with different area (A) to perimeter (P) ratio [4,5]. From the slope of the linear fit to the I/A versus P/A data, one can derive J_P and a similar procedure fitting I/P yields accurate J_A values.

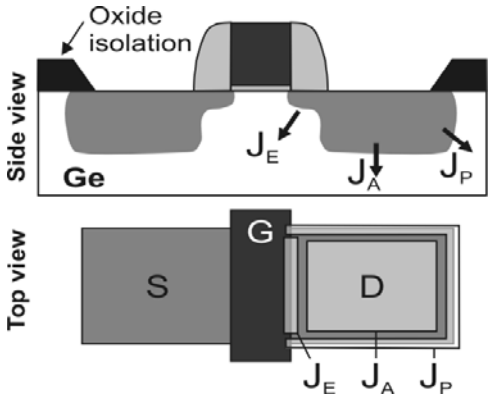


Figure 1. Cross-sectional (top) and birds-eye view of the drain-to-bulk leakage current components in a scaled MOSFET.

Ex. 1016, Page 89

The HDD and extension junctions in Ge MOSFETs are fabricated by means of ion implantation of p-type dopants in an n-well; when properly annealed to remove the implantation damage [1,5,6], J_P usually dominates J_A. Therefore, the aim of the present work is to further study the impact of the junction anneal temperature on the junction leakage current components and how this translates into the off-state transistor leakage current.

Experimental

Junctions and pMOSFETs have been processed in ~1.5 μm epitaxial germanium layers deposited on 200 mm Czochralski Si substrates by ASM. As the layer is fully relaxed, a threading dislocation density (TDD) of 10^8 cm^{-2} is present in the as-grown layer. After annealing the TDD is reduced to $2x10^7$ cm^{-2} which may impact J_A [7,8]. Standard n-well, halo and HDD implantations were implemented, as described in more detail in Ref. [9]. Two types of extensions have been evaluated: the standard one, using an 11 keV BF_2 implantation to a dose of $8x10^{14}$ cm^{-2} and a low-energy B implant following a Ge pre-amorphization (B+Ge in Table I). The main process variable was the junction annealing temperature, using Rapid Thermal Annealing (RTA), which, according to Table I was varied between 500 and 600°C. For the pMOSFETs, HfO_2 was deposited by Atomic Layer Deposition on a silicon passivated surface [10]. The thin epitaxial silicon layer was partially oxidized, using ozone chemistry in order to have a low density of interface states. A 10 nm TaN, 70 nm TiN metal gate was deposited by Physical Vapor Deposition, while source and drain were contacted by nickelgermanidation [9].

TABLE I. Processing splits for the Ge p^+n junctions and pMOSFETs

HfO_2 thickness (nm)	Extensions	RTA Temperature (°C)
2	BF_2	550
4	BF_2	500
4	Ge+B	500
4	BF_2	600

The *I-V* measurements of the p^+n junctions have been performed from 2 V to +1 V, with the bias applied to the n-substrate and the p^+ contact grounded. At least 5 dies per process condition have been measured, distributed evenly across the wafer, in order to have an idea about the reproducibility and spread of the results. Measurements have also been performed as a function of temperature (*T*) up to 140°C. Additionally, Capacitance-voltage (*C-V*) measurements at 1 MHz have been performed to extract the n-well doping density profile, the depletion region width W and the electric-field enhancement factor Γ, as outlined elsewhere [12]. The off-state source-to-drain current (I_D) of the pMOSFETs has been evaluated with 0 V gate voltage (V_{GS}).

Results and Discussion

First, the results of the p-n junction analysis will be reported, followed by the data for the Ge pMOSFETs, which will be compared with the former data.

Leakage Current of p-n Junctions

According to the J_A of Fig. 2, the best area leakage current density at 0.5 V is obtained for the Ge+B process split, while the RTA temperature does not have a particular impact. This implies that annealing at 500°C is sufficient to remove the HDD, halo and extension

Ex. 1016, Page 90

implantation damage [1,6,13,14]. In fact, it has been demonstrated that for an optimized processing, the area leakage current density can be accurately described in the frame of a trap-assisted-tunneling (TAT) model [15], based on the theory by Hurkx *et al.* [16,17]. This is illustrated by Fig. 3, showing a calculated fit to the experimental Γ for a 500°C annealed standard p^+n junction in the temperature range from 25 to 70°C. As shown elsewhere [12], an optimum fit is obtained for a reasonable tunneling effective mass of 0.15 m_0 (rest mass of the electron).

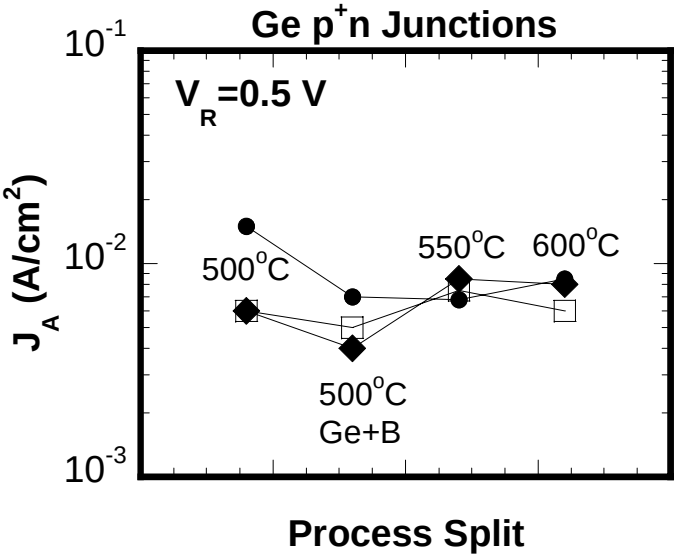


Figure 2. Area leakage current density at 0.5 V reverse bias and room temperature for Ge p^+n junctions using the process splits of Table I. Data on three dies is represented.

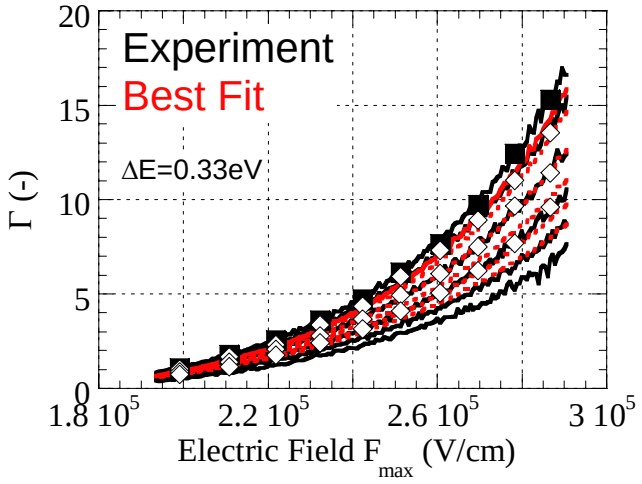


Figure 3. Best fit (lines) to the experimental Γ factor (symbols) using a barrier ΔE=0.33 eV, for the curves corresponding with T=25, 30, 40, 50, 60 and 70°C.

The situation is different with respect to J_P in Fig. 4, where a clear impact of the RTA temperature is observed. One can derive a significantly higher J_P at 0.5 V for the 550 and 600°C anneals, which is assigned to the higher activation of the P halo implantation. This results in a higher electric field near the Ge/SiO_2 interface and, hence, a higher field-assisted generation of leakage current. According to recent results on extension leakage, it could be due to enhanced Band-to-Band Tunneling (BTBT) in the peripheral region.

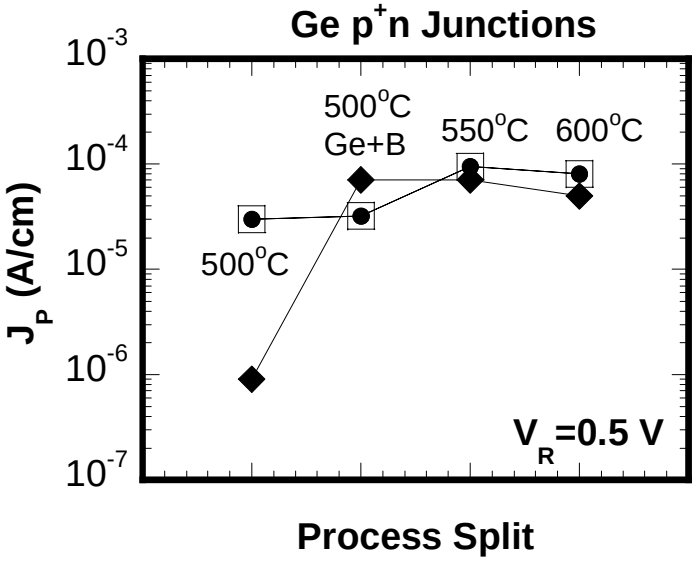


Figure 5. Perimeter leakage current density at 0.5 V reverse bias and room temperature for Ge p^+n junctions using the process splits of Table I. Data on three dies is represented.

Off-state Leakage of pMOSFETs

The off-state leakage current at V_{GS}=0 V and room temperature for the different process splits is shown in Fig. 5 for different channel lengths and a fixed mask width W=10 μm. The impact of the annealing temperature, the oxide thickness and the different type of extensions are shown. The 550°C anneal stands out as the most leaky one. From the comparison between the off-state leakage current and the perimeter current density in Fig. 6, it is obvious that they follow the same trend, with a maximum for the 550°C anneal. At the same time, the implementation of the Ge+B extension appears to have a beneficial effect on J_P and the transistor off-state leakage current. This could point to less residual implantation damage compared with the BF_2 extension case.

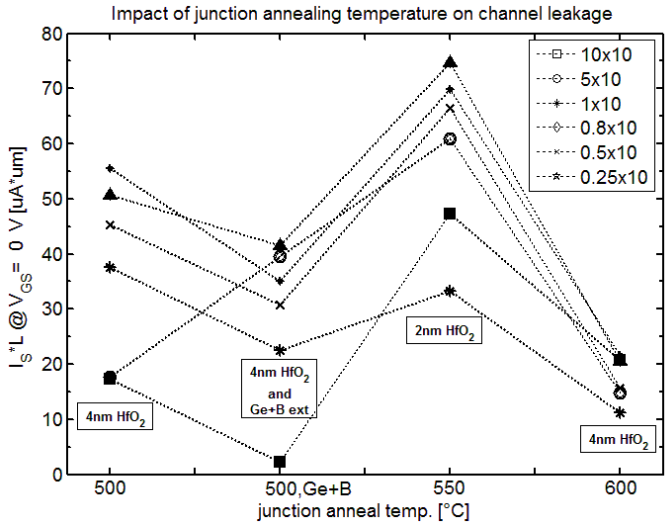


Figure 5. Off-state leakage versus junction anneal temperature and for different channel lengths. W=10 μm, V_{GS}=0 V and V_{DS}=-0.5 V.

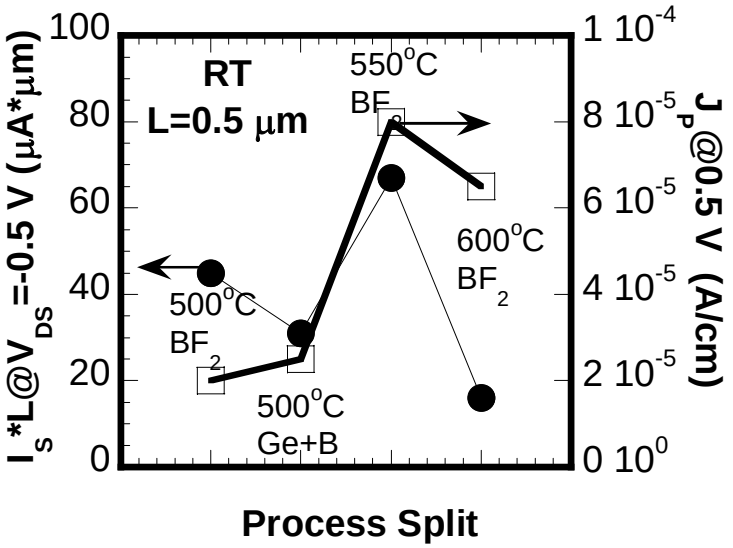


Figure 6. Comparison between the off-state leakage current I_Sx0.5 μm and the junction perimeter leakage current density J_P for the different process splits studied.

Conclusions

It has been shown that optimal Ge p^+n junction leakage current is obtained for a 500°C RTA temperature, in combination with a Ge pre-amorphization plus shallow B implantation extension. This also translates into a better off-state leakage current of the pMOSFETs made in the same substrates.

Acknowledgments

We would like to acknowledge the IMEC Ge/III-V Team for the processing of the devices.

References

1. E. Simoen. A. Satta, M. Meuris, T. Janssens, T. Clarysse, C. Demeurisse, B. Brijs, I. Hoflijk, W. Vandervorst and C. Claeys, *Solid State Phenomena*, 1**08-109**, 691 (2005).
2. J. Vanhellemont and E. Simoen, *J. Electrochem. Soc.*, **154**, H572 (2007).
3. G. Eneman, B. De Jaeger, G. Hellings, J. Mitard, E. Simoen, D.P. Brunco, K. De Meyer, M. Meuris and M.M. Heyns, submitted for publication in *IEEE Electron. Device Lett.*
4. E. Simoen, C. Claeys and J. Vanhellemont, *Defect and Diffusion Forum*, **261-262**, 1 (2007).
5. E. Simoen, S. Sonde, C. Claeys, A. Satta, B. De Jaeger, R. Todi and M. Meuris, *J. Electrochem. Soc*., **155**, H145 (2008).
6. A. Satta, G. Nicholas, E. Simoen, B. De Jaeger and J. Van Steenbergen, *Mater. Sci. in Semicond. Process*, **9**, 716 (2006).
7. E. Simoen, G. Brouwers, G. Eneman, M. Gonzalez Bargallo, F.E. Leys, B. De Jaeger, J. Mitard, D.P. Brunco, L. Souriau, N. Cody, S.G. Thomas, L. Lajaunie, M.-L. David and M. Meuris, To be published in: the *Proc. of Extended Defects in Semiconductors,* Poitiers, 14-19 Sept., 2008, *Phys. Stat. Sol*.

8. E. Simoen, G. Eneman, P. Verheyen, R. Loo, M. Bargallo Gonzalez and C. Claeys, in: the *Proc. of the ECS Fall Symp. on SiGe, Ge & related Compounds: Materials, Processing, and Devices*, ECS Trans., **16**(10), p. 513 (2008).
9. G. Nicholas, B. De Jaeger, D.P. Brunco, P. Zimmerman, G. Eneman, K. Martens, M. Meuris and M.M. Heyns, *IEEE Trans. Electron Devices*, **54**, 2503 (2007).
10. F.E. Leys, R. Bonzom, B. Kaczer, T. Janssens, W. Vandervorst, B. De Jaeger, J. Van Steenbergen, K. Martens, D. Hellin, J. Rip, G. Dilliway, A. Delabie, P. Zimmerman, M. Houssa, A. Theuwis, R. Loo, M. Meuris, M. Caymax and M.M. Heyns, *Mater. Sci. Semicond. Process.*, **9**, 679 (2006).
11. D. Brunco, B. De Jaeger, G. Eneman, J. Mitard, G. Hellings, A. Satta, V. Terzieva, L. Souriau, F. Leys, G. Pourtois, M. Houssa, G. Winderickx, E. Vrancken, S. Sioncke, K. Opsomer, G. Nicholas, M. Caymax, A. Stesmans, J. Van Steenbergen, P. Mertens, M. Meuris and M. Heyns, *J. Electrochem. Soc*., **155**, H552 (2008).
12. E. Simoen, F. De Stefano, G. Eneman, B. De Jaeger, C. Claeys and F. Crupi, submitted for publication in *IEEE Electron Device Lett*.
13. A. Satta, E. Simoen, T. Clarysse, T. Janssens, A. Benedetti, B. De Jaeger, M. Meuris and W. Vandervorst, *Appl. Phys. Lett*., **87**, 172109-1 (2005).
14. A. Satta, E. Simoen, T. Janssens, T. Clarysse, B. De Jaeger, A. Benedetti, I. Hoflijk, B. Brijs, M. Meuris and W. Vandervorst, *J. Electrochem. Soc*., **153**, G229 (2006).
15. G. Eneman, M. Wiot, A. Brugère, D. Brunco, B. De Jaegere, A. Satta, O. Sicart i Castain, S. Sonde, C. Claeys, M. Meuris and E. Simoen, *IEEE Trans. Electron Devices*, **55**, 2287 (2008).
16. G.A.M. Hurkx, D.B.M. Klaassen and M.P.G. Knuvers, *IEEE Trans. Electron Devices*, **39**, 331 (1992).
17. G.A.M. Hurkx, H.C. de Graaff, W.J. Kloosterman and M.P.G. Knuvers, *IEEE Trans. Electron Devices*, **39**, 2090 (1992).

ECS Transactions, 18 (1) 67-70 (2009)
10.1149/1.3096429 © The Electrochemical Society

A Study of Stress technology for 65nm Generic CMOS technology

Xinfu Liu[a], Wenjun Dai[b],Xing Yu[a], Louis Lim[a], Bin Xue[a], Boyong He[a]

[a] Department of ACT, Chartered Semiconductor Manufacturing LTD, 738406, Singapore
[b] Department of ACE, United Microelectronics Corporation, 519528, Singapore

Abstract

For deep sub-100nm CMOS technology, the stress level impacts transistor performance significantly. In this paper, we present the investigation results of stress technology in a 65 nm generic process. We studied some mobility enhancing technologies such as Stress-Memorization Technology (SMT) and Dual Stress Liner (DSL) process in a conventional 65 nm Generic CMOS process. The transistor performance with different stress layers and stress levels are reported. The experimental results show that SMT and higher stress level tensile ESL (Etch-Stop Layer) improve NMOS performance while compressive ESL improve PMOS significantly. For the required-device performance, the stress layer and stress architecture should be carefully optimized. The stress technology reported in this paper has been successfully implemented in the production of 65 nm generic process.

Introduction

Explosive demand for wireless, digital consumer, and mobile products, has driven system-on-a-chip (SoC) requirements for density, performance and cost and caused very hard competition of technology development, which demands us to enhance the chip performance and reduce the power consumption. The scaling of transistor dimension and operation voltage are the conventional approaches to achieve the requirement. As gate lengths approach sub-65 nm dimensions and gate oxides approach around 1 nm, scaling becomes more challenging. Fundamental physical limitations are difficult to overcome by traditional semiconductor technology and materials. Moreover, transistor scaling down, higher channel doping concentrations and more abrupt, shallow source-drain junctions used to control short-channel effects at very short gate lengths result in carrier mobility degradation, increasing threshold voltage variation, junctions leakage and capacitance. To improve device performances, much attention has been focused on the high-mobility channel engineering that exhibits increased inversion layer mobility and higher carrier velocities in short channel devices by stress technology.

The usage of stress technology to enhance CMOS performance, however, will increase process complexity and cost. For high volume mass production, the stress technology needs to balance the device performance and cost.

In this paper we report the stress technology and results in our 65 nm generic process for high-performance application.

Experiments and results

Ex. 1016, Page 95

The transistor were fabricated in a 65 nm generic process platform using (100) substrate. The transistor channel is (110) by rotating the wafer 45 degree. A tensile nitride layer is deposited on the NMOS area followed by SPIKE anneal to boost NMOS mobility, then this nitride layer is removed by wet strip. This technology is so-called Stress-Memorization Technology (SMT). Compared to low power process which uses single tensile Contact-Etch-Stop Layer (ESL), dual stress layer (DSL) for 65 generic process was studied to enhance NMOS and PMOS performance simultaneously. Tensile nitride layer is used as the ESL on NMOS which enhances the NMOS device performance, while a compressive nitride layer is used to enhance PMOS performance.

Different stress experiments to study the device performance enhancement are shown in table 1.

Table1 . Experiment Plan for Stress

No	Experiment Plan	Results
1	60nm tensile ESL	Tensile ESL Improves NMOS, and if tensile ESL is used, PMOS is not sensitive to the thickness of tensile ESL and stress level
2	90 nm tensile ESL	
3	0.7GPa tensile ESL	
4	1.2 GPa tensile ESL	
5	SMT only for NMOS	(1)SMT significantly improves NMOS and degrades PMOS (2) Tensile SMT significantly degrades PMOS but the stress can be relaxed by Ge implantation
6	NMOS skip SMT	
7	PMOS with tensile SMT	
8	PMOS with tensile SMT but released by Ge implantation	
9	50 nm tensile SiN film for NMOS SMT	Different thickness of SMT layer does not impact the device performance for both NMOS and PMOS
10	75 nm tensile SiN film for NMOS SMT	
11	Compressive ESL for PMOS	Different type of ESL impacts the transistor performance. Tensile ESL improves NMOS but degrades PMOS. While compressive ESL improves PMOS but degrades NMOS
12	Tensile ESL for PMOS	

Fig.1 and Fig.2 show the NMOS and PMOS Idsat_Ioff curve with different thickness of ESL. For ESL thickness experiments, 60 nm and 90 nm tensile ESL were used. It is very clear that the NMOS performance is sensitive to ESL thickness, but PMOS performances are comparable for different thickness of ESL. This is because the channel of transistor is (110) and PMOS performance is not sensitive to small range of stress difference.

Fig.3 and Fig.4 show the NMOS and PMOS performance difference with different stress level of ESL. Similar to Fig.1 and Fig.2, the NMOS is sensitive to stress level, and NMOS improves 5% when the stress level increases to 1.2GPa from 0.7 GPa. But PMOS are comparable for different tensile level stress between 0.7GPa and 1.2GPa.

For NMOS performance improvement, 1.2GPa tensile nitride with 50 nm thickness is used followed by RTA. This layer is so-called SMT layer. Fig.5 shows the device performance with and without tensile SMT nitride. It shows that the NMOS performance is improved 10% with SMT. Fig.6 shows the PMOS difference if the transistor follows same process as NMOS. The PMOS degrades more than 20% if it follows tensile SMT. But when we add one step of Ge implantation to the SMT layer on PMOS to relax the

tensile stress, the PMOS device performance will be recovered by 15% , but it is still worse than no-SMT process because the tensile stress can not been fully relaxed.

Fig.7 and Fig.8 show the results for different thickness of SMT nitride. There is no difference between 50 nm and 75 nm tensile SMT nitride layer for both NMOS and PMOS. That means that the stress levels are saturated even if we change the thickness of SMT. The SMT stress is memorized in Poly gate after RTA and changes the channel stress level under the Poly gate, the Poly grain size and thickness possibly are the key factors of stress transition from nitride to channel, and in some range of Poly grain size and Poly thickness, the stress transition rate will be saturated.

Fig.9 and Fig.10 show the results with different types of ESL. Tensile ESL and Compressive ESL were studied to compare their impact on NMOS and PMOS respectively. Before ESL process, the NMOS run through the SMT process while the PMOS skip SMT.

Fig.9 shows that there is 5% difference between tensile and compressive ESL for NMOS. Tensile improves NMOS while compressive degrades NMOS. Because NMOS has been processed with SMT already, the difference is not significant for different types of ESL. In Fig.10, we can find more than 10% difference for different type of ESL for PMOS. Because stress level and stress type between compressive ESL and tensile ESL is significant and there is no SMT for PMOS, the PMOS device shows much performance difference for different type of ESL than NMOS.

From the results shown above, we optimize the stress technology in our 65 nm generic platform and successfully implement them in the production of 65 nm generic products. Fig.11 and Fig.12 shows the TEM of a NMOS and PMOS nominal transistor of the 65 nm generic process.

Conclusion

For the 65 nm generic process, the stress technology is very critical to balance cost and performance in volume production. In this paper, we studied stress factors of SMT and different ESL parameters such as film type, stress level and film thickness. The device performances with different stress processes are compared. Tensile ESL and SMT improve NMOS device performance but degrade PMOS significantly, and compressive ESL improves PMOS but degrades NMOS. The combination of tensile ESL with SMT for NMOS and compressive ESL for PMOS is chosen for the 65 nm generic process to achieve satisfied NMOS and PMOS performance.

Reference

[1] Samuel K.H. Fung, et al. *65 nm CMOS High Speed, General Purpose and Low Power Transistor Technology for High Volume Foundry Application*, 2004 symposium on VLSI Technology Digest of Technical Papers.

[2] S.Saito, *Orgin of Drivability Enhancement in Scaled pMOSFET with 45degree Rotated <100> channels*, 2006 symposium on VLSI Technology Digest of Technical Papers.

Ex. 1016, Page 97

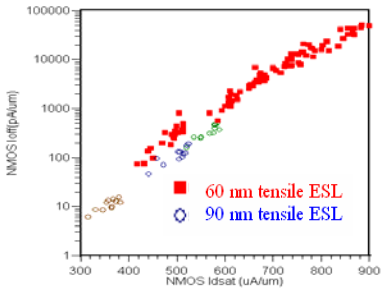


Fig.1. NMOS Idsat_Ioff with Different Thickness of Tensile ESL

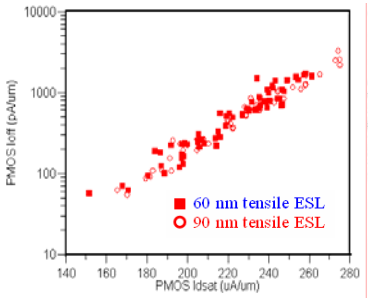


Fig.2. PMOS Idsat_Ioff with Different Thickness of Tensile ESL

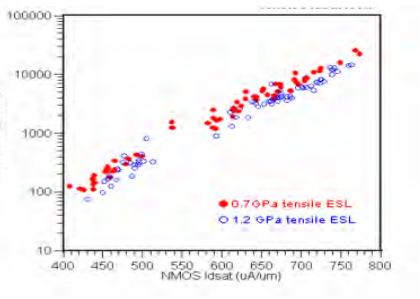


Fig.3. NMOS Idsat_Ioff with Different Stress Level ESL

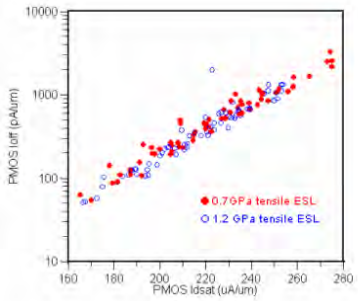


Fig.4. PMOS Idsat_Ioff with Different Stress Llevel ESL

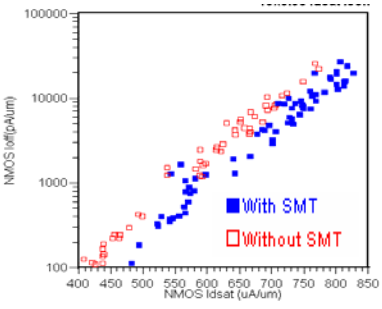


Fig.5. NMOS Idsat_Ioff for SMT Experiment

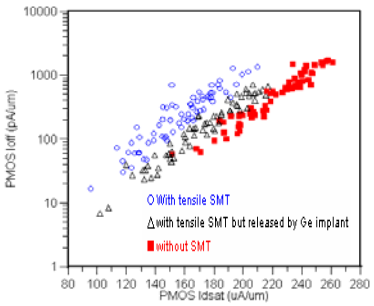


Fig.6. PMOS Idsat_Ioff for SMT Experiment

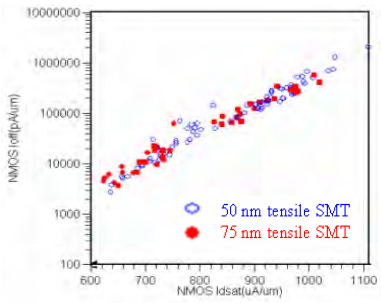


Fig.7. NMOS Idsat_Ioff for SMT Thickness Experiment

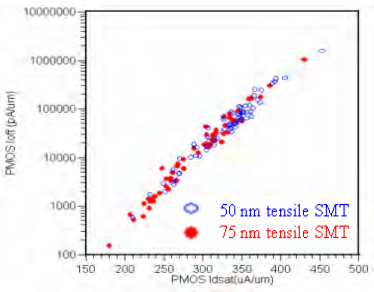


Fig.8. PMOS Idsat_Ioff for SMT Thickness Experiment

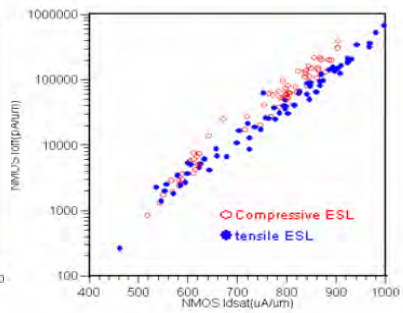


Fig.9. NMOS Idsat_Ioff with Different type of Stress ESL

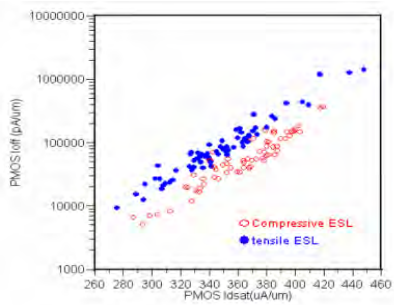


Fig.10 .PMOS Idsat_Ioff with Different Type of Stress ESL

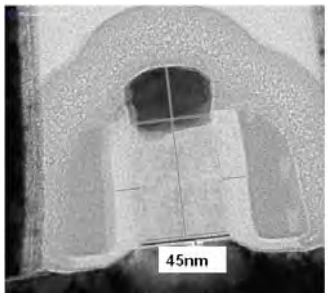


Fig.11 NMOS of TEM of a 65 nm Generic Process

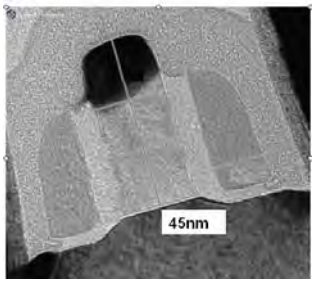


Fig.12 PMOS of TEM of a 65 nm Generic Process

ECS Transactions, 18 (1) 71-75 (2009)
10.1149/1.3096430 © The Electrochemical Society

Evaluation of Lateral Ni Diffusion in Si Nanowire Schottky Contact

H. Kamimura[a], H. Arai[a], S. Sato[a], K. Kakushima[b],
P. Ahmet[b], K. Tsutsui[b], N. Sugii[b], T. Hattori[a] and H. Iwai[a]

[a]Frontier Research Center, Tokyo Institute of Technology,
[b]Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology
4259 S2-20, Nagatsuta, Midori-ku, Yokohama, 226-8503, Japan

Lateral Ni diffusion in Si Nanowire (SiNW) to form NiSi Schottky contact was characterized by scanning electron microscope at temperatures range from 400 to 700 ^{o}C. Higher temperature annealing results in longer NiSi formation and the length of NiSi increases with the annealing time. The activation energy for NiSi formation on SiNWs are found to be 0.9 eV which is lower than that on bulk case. One-dimensional diffusion of Ni with SiNW might be the reason for lowering the activation energy.

Introduction

Si Nanowire field effect transistors (SiNW-FET) have attracted much attention for future device, owing to its ability to suppress the off-state leakage current. One of the concerns in SiNW-FET is the increase in access resistance at source/drain region, which eventually reduces the on-state current. Commonly, metal silicide contacts, especially NiSi, have been used to reduce the resistance for bulk and SOI FET. However, the lateral diffusion of Ni to form NiSi in SiNW is still unknown (1-4). In this paper, evaluation of lateral Ni diffusion into SiNW has been conducted.

Experimental

Si lines with different widths ranging from 60 to 100 nm along <100> were formed on a SOI substrate (SOI and BOX thickness of 61 and 145 nm, respectively) by dry etching. Wafer was then subjected to thermal oxidation in dry O_2 at 1100 ^{o}C for 30min, resulting in formation of SiNWs surrounded by SiO_2. After striping a part of the formed SiO_2 with buffered HF to expose SiNWs, 10-nm-thick Ni film was deposited by magnetron sputtering in Ar gas at a pressure of 5.5×10^{-1} Pa. Silicidation annealing was performed by rapid thermal annealing (RTA) apparatus in a forming gas (N_2/H_2=97%/3%) ambient at annealing temperatures range from 400 to 700 ^{o}C . The unreacted Ni was removed by H_2SO_4/H_2O_2 mixture. The formed NiSi was observed and measured by field-emission scanning electron microscope (SEM) at a high acceleration voltage (10 keV) so that the secondary electrons from NiSi can be collected through 50-nm-thick SiO_2 shell.

Ex. 1016, Page 99

Results and Discussion

Figure 1 shows a typical SEM image of a lateral NiSi formation in a SiNW. This sample was fabricated at an annealing temperature of 450 ^{o}C for 30 sec. A bright contrast in the SiNW indicates the formation of NiSi by lateral Ni diffusion. The length (λ) of the formed NiSi wire was measured from the edge of the SiNW.

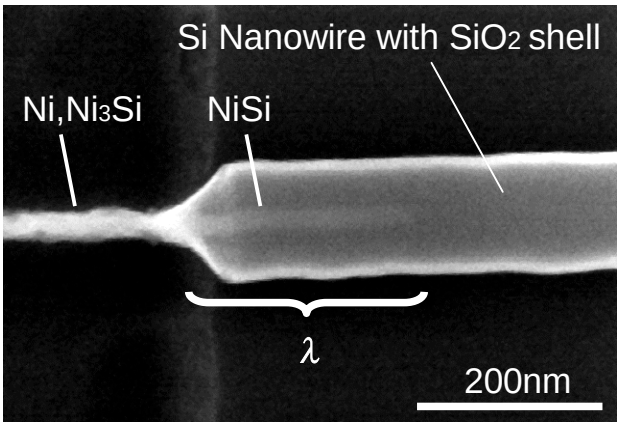


Figure 1. SEM image of NiSi formed by lateral Ni diffusion into SiNW.

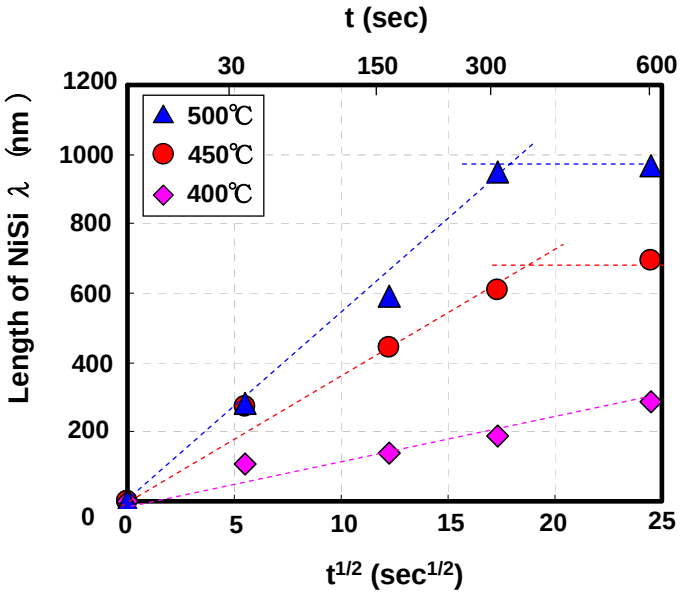


Figure 2. Length of NiSi on square root of the annealing time.

Figure 2 shows the length of NiSi formed at 400, 450 and 500 ^{o}C on a SiNW with a cross section of 7.5 x 12 nm^2. Higher temperature annealing results in longer NiSi formation and the length of NiSi increases with the annealing time. Therefore, it can be speculated that NiSi formation on SiNW is a thermally activated process. However, after annealing for 300 sec, the formation seems to saturate.

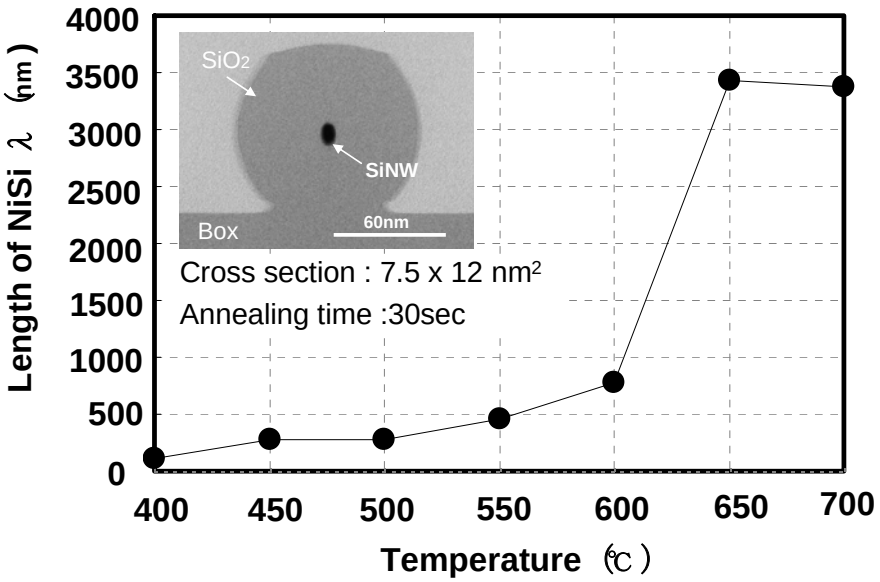


Figure 3. Diffusion length of Ni versus temperature range 400 – 700 °C.

The dependence of the length of NiSi on annealing temperature is shown in figure 3. A monotonic relation was observed up to a temperature of 600 °C, however, a progressive formation over 3 μm was observed above 650 °C. Generally, the temperature required for phase transition from NiSi to $NiSi_2$ for thin film case is known to be 700 °C (5). As the cross section is nanometer size and also that the direction of Ni diffusion is one-dimension for our experiment, a lowering of phase transition temperature down to 650 °C might happen. Figure 4(a) and (b) show SEM images of a sample annealed at 650 °C for 30 sec at the edge of exposed region. Grains with a size of 80 nm were observed at the entrance of NiSi wire. These grains were also observed at a temperature of 500 °C for 600 sec although the size is much smaller. Considering that there was almost no volume change by NiSi formation in SiNW, the lateral diffusion of Ni may be accompanied with a lateral diffusion of Si in opposite direction. As these grains were observed at the entrance of NiSi, the Si atoms may diffuse along $NiSi/SiO_2$. Therefore, it can be speculated that both Ni and Si are the diffusing species for NiSi formation in SiNW.

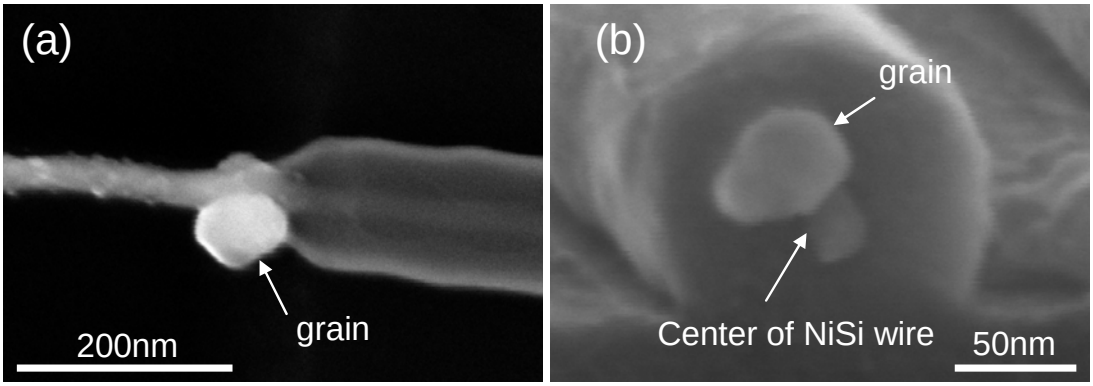


Figure 4. SEM image of (a) top view and (b) cross section of Ni diffusion in SiNW after annealing at 650 °C for 30 sec.

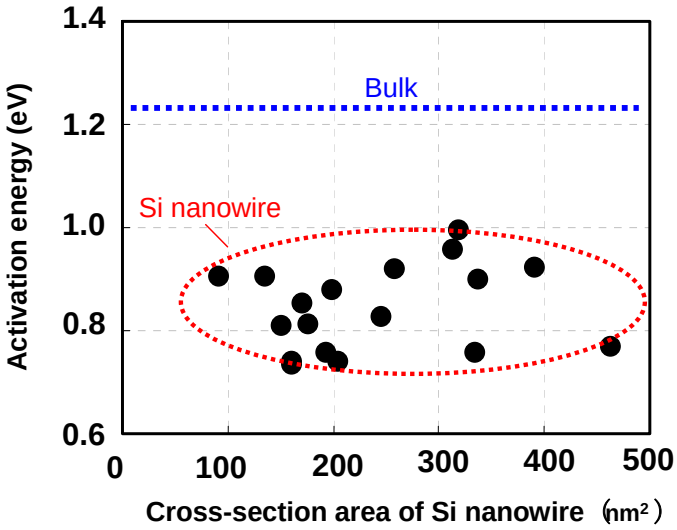


Figure 5. Plot of activation energy versus cross-section area of SiNW.

The activation energy dependence of NiSi formation to SiNW on the cross sectional areas is shown figure 5. An average activation energy of 0.9 eV has been obtained irrespective to the cross sectional area. The activation energies of the Ni diffusion in <100> SiNWs are't reacted on the cross-section area of SiNW. As the activation energy for NiSi formation on bulk Si(100) wafer is 1.2 ± 0.3 eV (6), the smaller activation energy for SiNW might come from one dimensional direction during the diffusion.

Conclusion

Lateral Ni diffusion in Si Nanowire to form NiSi Schottky contact was characterized by scanning electron microscope at temperatures range from 400 to 700 ^{o}C. Higher temperature annealing results in longer NiSi formation and the length of NiSi increases with the annealing time. The activation energy for NiSi formation on SiNWs are found to be 0.9 eV which is lower than that on bulk case. One-dimensional diffusion of Ni with SiNW might be the reason for lowering the activation energy.

Acknowledgments

This study has been supported by METI's Innovation Research Project on Nanoelectronics Materials and Structures.

References

1. C. A. Decker, R. Solanki, J. L. Freeouf, and J. R. Carruthers, Applied Physics Letters, vol. 84, no. 8, pp.1389-1391 (2004).

2. J. Appenzeller, J. Knoch, E. Tutuc, M. Reuter, and S. Guha , Technical Digest of IEDM, pp. 555 (2006).
3. J. T. Sheu, S. P. Yeh, C. H. Lien, and S. T. Tsai, Japanese Journal of Applied Physics, vol. 45 , no. 4B, pp. 3686-3689 (2006).
4. Kuo-Change Lu and K.N.Tu, Applied Physics letters, vol 90, 253111 (2007).
5. F. F. Zhao, J. Z. Zheng, Z. X. Shen, T. Osipowicz, W. Z. Gao, L. H. Chan, Microelectronic Engineering, vol. 71, pp. 04-111 (2004).
6. G. Majni, F. Della Valle and C. Nobili, Journal of Physics D: Appl. Phys, vol. 17, pp.L77-L81 (1984).

ECS Transactions, 18 (1) 77-81 (2009)
10.1149/1.3096431 © The Electrochemical Society

Characteristic Variations Induced by Random Fluctuation of Source/Drain Lateral Distribution Dopant in Nano-Scale UTB SOI MOSFET

Linfeng Du[a], Hui Deng[a], Gang Du[a], Ruqi Han[a] and Shengdong Zhang[b]

[a] Institute of Microelectronics, Peking University, Beijing 100871, PRC
[b] Shenzhen Graduate School, Peking University, Shenzhen 518055, PRC

The "nominally" un-doped or lightly-doped channels are actually doped considerably due to the lateral distribution of dopants from Source/Drain (S/D) doping, thus forming a new source of Random Dopant Fluctuation (RDF). In this work, the unexpected RDF effect of sub-20nm UTB SOI MOSFETs is investigated by device simulation. A new method to simulate the RDF effect was introduced. The results show that a severe V_{th} variation due to the new RDF effect will take place when device is scaled into sub-20 nm regime, even with un-doped channel. The device parameter dependence of V_{th} variations of MOSFETs with un-doped channel were presented, which was found to be largely similar to that of ones with doped channel. The drain current and subthreshold fluctuations were also investigated to gain a more comprehensive insight to RDF in MOSFETs. Our results illustrated the importance of suppressing the lateral distribution of S/D dopants for scaled devices.

I. Introduction

The random dopant induced characteristic fluctuation is one of fundamental challenges in shrinking nano-scaled scaling devices (1). The un-doped or lightly-doped ultra-thin-body (UTB) is generally recommended for the ultra-small MOSFETs. Moreover, the UTB devices have superior electrostatic integrity for short gate length devices compared with halo-doped bulk silicon or partially depleted SOI MOSFETs (2), (3). However, lateral distribution dopant from source/drain (S/D) will intensely influence the effective doping level in the channel, especially in extremely scaled devices, thus forming the new sources of RDF. The S/D lateral distribution is fundamentally inevitable after S/D ion implantation, diffusion, and thermal annealing. Therefore, it is necessary to consider the unintended dopant due to the S/D lateral distribution when one evaluates the immunity of MOSFETs to RDF. In this work, we investigate characteristic fluctuations in scaled UTB SOI MOSFETs with "nominally" un-doped body. This paper is organized as follows. In section II, the method for analyzing the RDF effects in nanoscale device is described in detail. In section III, we present and discuss discrete dopant induced characteristic fluctuations of nanoscale transistor. Finally conclusions are drawn.

II. Simulation Method

In this section, a new method to introduce RDF in "nominally" un-doped channel was presented. Figure 1 shows the device structure of UTB SOI MOSFET assumed in the

Ex. 1016, Page 105

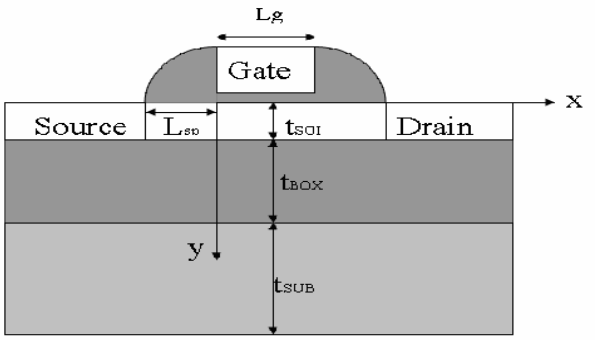


Figure 1. Device structure of UTB SOI MOSFET assumed in the study.

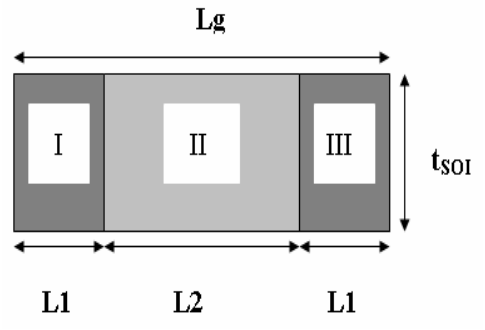


Figure 2. The channel is divided into three regions.

TABLE I. DEVICE PARAMETERS OF UTB SOI MOSFET USED IN THE SIMULATION					
t_{BOX}	W_g	t_{SUB}	L_{sp}	N_{SUB}	N_{poly}
40nm	50nm	50nm	15nm	10^{15} cm^{-3}	10^{20} cm^{-3}

TABLE II. ION IMPLANTATION CONDITIONS FOR S/D				
Species	**Scattering oxide thickness**	**Implantation Energy**	**Dose**	**Annealing time**
P	6nm	8keV	$2.1x10^{14}cm^{-2}$	1S

study. The device parameters used in the device simulation are shown in Table I. The SOI thickness (t_{SOI}), the gate length (L_g), and the gate oxide thickness (t_{OX}) are variable parameters in the device simulation. The S/D doping concentration ($N_{S/D}$), the S/D extension doping concentration (N_{ex}) and the distribution of the induced channel impurity are entirely obtained by using a 2-D process simulator DIOS (4). The ion implantation conditions for S/D assumed in our work are summarized in Table II. The implantation energy is chosen so that the implanted ions can be effectively injected into S/D without going through the buried oxide. The implantation dosage of phosphorus ion is chosen so that $N_{S/D}$ is almost equal to 2 x 10^{20} cm^{-3} after one second annealing. The DIOS simulated $N_{S/D}$, N_{ex} are 2.26 x 10^{20} cm^{-3} and 7.1 x 10^{19} cm^{-3}, respectively, which are assumed in device simulation.

The channel discrete dopants arising from S/D lateral distribution are mostly located in regions near S/D. The middle region of the channel is lightly-doped. Therefore, the channel is divided into three regions (I, II, III). Figure 2 shows how the channel is divided. In order to decide the dimension of every region, one could notice that most discrete dopants are located in I and III regions. By assuming that one dopant is located in the middle region (II), one can use equation [1].

$$W_g L_2 t_{SOI} N_2 = 1 \qquad [1]$$

where N_2, L_2 are the average dopant concentration and the length of the region II respectively, W_g is the channel width of MOSFETs. L_2 is chosen so that equation [1] is approximately true, thus almost including one dopant in region II. L_1 is computed by using $(L_g-L_2)/2$. One can use equation [2] to compute the dopant number in region I/III, where N_1 is the average concentration of the other two regions (I, III). N_1, N_2 are obtained by using process simulator DIOS (4). We had computed the dopant numbers in

Ex. 1016, Page 106

region I/III (n_1) for devices of different structure parameters, such as L_g, t_{si}. For the MOSFET where L_g = 20 nm, W_g = 50 nm, the computed n_1 were 6 and 3 for t_{si}=10 and 6 nm, respectively. The divided three regions are so small that the particular distribution of dopants in these regions is not so obvious. Therefore, we assumed that dopants in these regions follow a uniform distribution respectively (5).

$$W_g L_1 t_{SOI} N_1 = n_1 \quad [2]$$

In order to introduce the RDF effects into the device simulation, one can use equation [3] (6) in the three regions respectively, where n is the dopant number of one region, $\sigma(n)$ is the standard deviation of the dopant number in the region.

$$\sigma(n) = \sqrt{n} \quad [3]$$

For carrier transport, the hydrodynamic model is adopted with the van Dort quantum correction. The doping dependent Mobility Model, the HighFieldSaturation Mobility Model, and the Enormal Mobility Model are included. For carrier recombination, the SRH and Avalanche Model are adopted. The device simulation is carried on with simulator DESSIS (4).

III. Simulation results and discussion

In this section, we give a insight into the effect of RDF on device characteristics by investigating hundreds of microscopically different MOSFETs. Figure 3 shows the simulated V_{th} distribution for MOSFETs with L_g= 20 nm, t_{si}=10 nm, and t_{ox}=1.2 nm. V_{th} is extracted with a current criteria 10^{-7} W_g/L_g (A) (at V_{ds}=0.05 V). 512 cases are assumed to compute the frequency distribution. The random dopant induced V_{th} variation almost follows a normal distribution. The standard deviation of V_{th} is 35.7 mV for these MOSFETs.

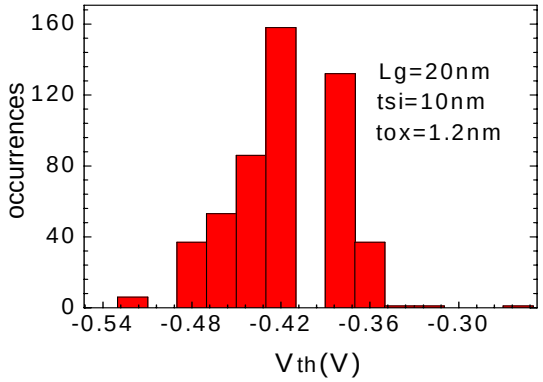


Figure 3. Frequency distribution of V_{th} for 512 microscopically different MOSFETs.

The V_{th} fluctuations of MOSFETs with "nominally" un-doped channel illustrated a similar dependence on the gate oxide thickness and the gate length to ones with doped channel (7). A sample of 100 microscopically different MOSFETs was assumed. The standard deviation of V_{th} versus the gate oxide thickness is shown in Figure 4. The V_{th} variation increases with increasing t_{ox}. The V_{th} variation of MOSFETs of L_g=16 nm, t_{si}=8

Ex. 1016, Page 107

nm is larger than that of L_g=20 nm, t_{si}=10 nm. It is concluded that smaller devices are more likely to be susceptible to RDF. Figure 5 shows the standard deviation of V_{th} versus the gate length. A sample of 100 microscopically different MOSFETs was assumed. It is clearly shown that the V_{th} fluctuation increases with decreasing L_g. The V_{th} variation of MOSFETs of t_{si}=6nm is less than that of t_{si}=10nm. Therefore, one can assume thinner silicon body to reduce the V_{th} fluctuation.

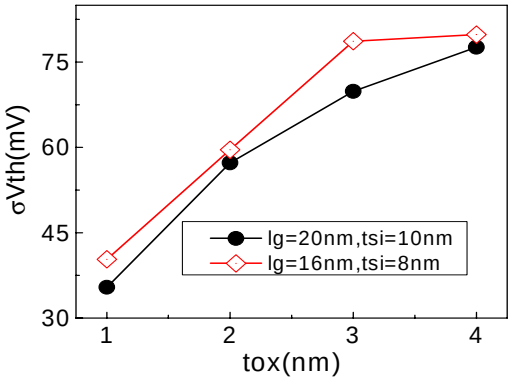


Figure 4. Standard deviation of V_{th} versus the gate oxide thickness (t_{ox}).

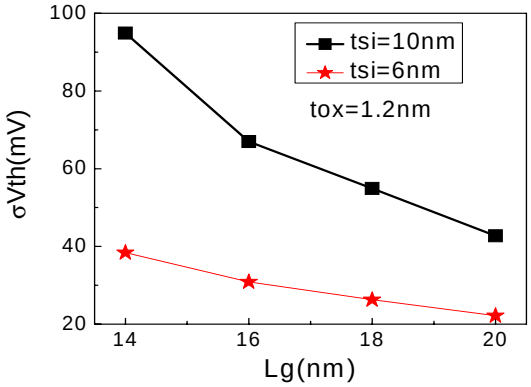


Figure 5. Standard deviation of V_{th} versus the gate length (L_g), t_{ox}=1.2 nm.

Figure 6 shows the simulated on-state current (I_{on}) versus off-state current (I_{off}). It is concluded that the RDF induced current variation is more distributive with increasing t_{si} and decreasing L_g, because the V_{th} variation increase for larger t_{si} and smaller L_g. We extracted I_{on} at V_{ds}=1 V, V_g=0.5 V and I_{off} at V_{ds}=1 V, V_g= -0.5 V (8). A sample of 100 microscopically different MOSFETs was assumed.

The frequency distribution of subthreshold slop (S) for MOSFETs of L_g=20 nm, t_{ox}=1.2 nm is shown in Figure 7. 100 microscopically different MOSFETs were assumed. The average subthreshold swing of MOSFETs of t_{si}= 5 and 10 nm are 90.56 and 168.7mV/dec respectively. The S histogram is more distributive for devices with thicker silicon body. It is concluded that thinner silicon body can lower the value of subthreshold swing and its fluctuation for MOSFETs with un-doped channel. Figure 8 illustrates the fluctuation in I_d-V_g characteristics for MOSFETs of L_g=20 nm, t_{si}=10 nm and t_{ox}=1.2 nm. A sample of 100 microscopically different MOSFETs was assumed. The RDF induced I_d-V_g characteristic fluctuation is more obvious in the subthreshold regime. These curves gradually get coincident with increasing V_g.

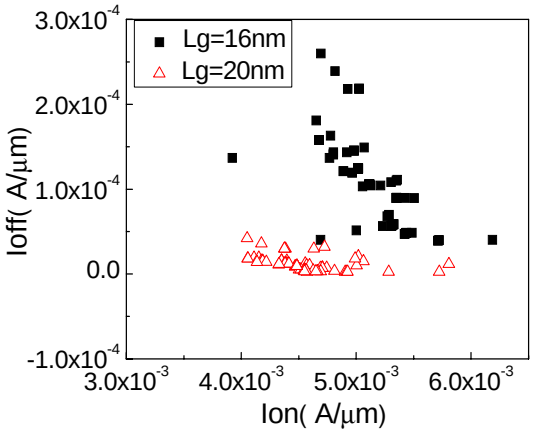


Figure 6(a). The simulated I_{off} versus I_{on}. t_{si}=6 nm, t_{ox}=1.2 nm.

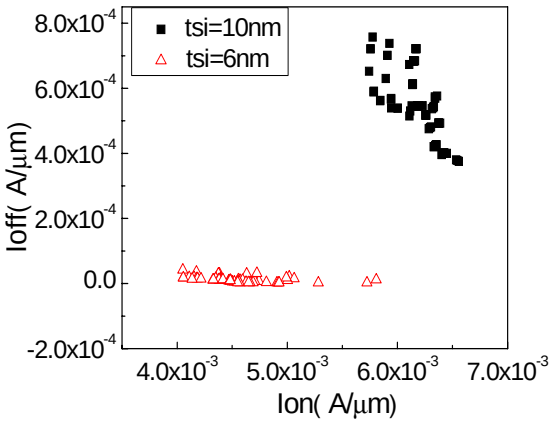


Figure 6(b). The simulated I_{off} versus I_{on}. L_g=20 nm, t_{ox}=1.2 nm.

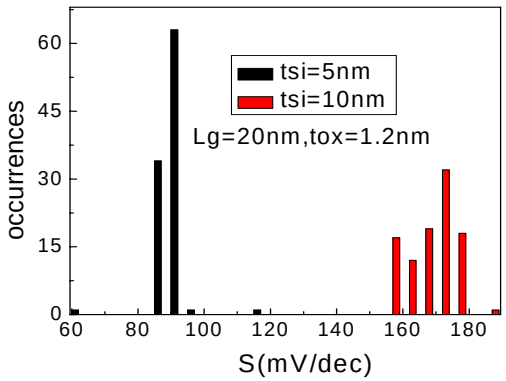


Figure 7. Frequency distribution of sub-threshold slop, L_g=20 nm, t_{ox}=1.2 nm.

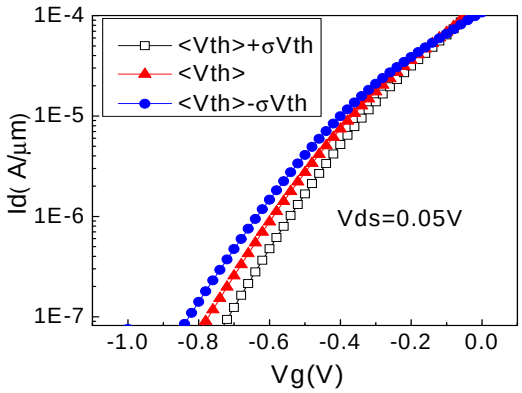


Figure 8. I_d-V_g characteristics at the $<V_{th}>$, $<V_{th}> \pm \sigma V_{th}$ points, V_{ds}=0.05V.

It is concluded that MOSFETs with "nominally" un-doped channel are actually susceptible to RDF effects due to the lateral distribution of S/D dopants. Suppressing the lateral distribution of S/D dopants is crucial for the immunity to RDF for scaled devices with un-doped channel.

IV. Conclusion

The characteristic variations of UTB SOI MOSFETs with "nominally" un-doped channel are investigated by device simulations. When the S/D lateral distribution is taken into account, UTB SOI MOSFETs with extremely scaled size were shown to be susceptible to RDF, even with un-doped channel. A new method to simulate the statistical dopant variations was described in detail. The V_{th} fluctuations of MOSFETs with un-doped channel illustrated a similar dependence on the gate oxide thickness and the gate length to ones with doped channel. In order to be immune to RDF, it is necessary to suppressing the lateral distribution of S/D dopants for deep nano-scaled MOSFETs.

Acknowledgement

This work is supported by NSFC of China under Project 60676023 and 60736030, and a foundation for the author of national excellent doctoral dissertation (FANEDD) of China under Project 200440.

References

1. D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H. S. P. Wong, *proc. IEEE*, **89**, 259-285 (2001).
2. K. Suzuki, T. Tanaka, Y. Tosaka, H. Horie, and Y. Arimoto, *IEEE Trans. Electron Devices*, **40**, 2326–2329 (1993).
3. J. Kedzierski, M. Ieong, T. Kanarsky, Y. Zhang, and H. S. P. Wong, *IEEE Trans. Electron Devices*, **51**, 2115–2120 (2004).
4. *User's manual for ISE TCAD Release 10.0.*
5. Y. Li, C.-H. Hwang, and H.-M. Huang, *phys.stat.sol.(a)* **205**, 1509 (2008).
6. T. Mizuno, J.-i. Okamura, and A. Toriumi, *IEEE Trans. Electron Devices*, **41**, 2219 (1994).
7. A. Asenov, *IEEE Trans. Electron Devices*, **45**, 2508-2510 (1998).
8. A. Majumdar *et al.*, *IEEE Electron Device Lett.*, **29**, 517 (2008).

ECS Transactions, 18 (1) 83-88 (2009)
10.1149/1.3096432 © The Electrochemical Society

A Novel Accumulation Mode GAAC FinFET Transistor: Device Analysis, 3D TCAD Simulation and Fabrication

Deyuan Xiao[a b], MinHwa Chi[b], David Yuan[c], Xi Wang[a], Yuehui Yu[a], Hanming Wu[b], Joseph Xie[b]

[a] Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050
[b] Semiconductor Manufacturing International Corporation, Shanghai 201203
[c] Synopsys Inc., Shanghai 200050

We report the analysis and TCAD results on gate-all-around cylindrical (GAAC) FinFETs with operation based on channel accumulation. GAAC FinFETs operating in accumulation mode in which the current in accumulation is flowing across the whole cylinder-shaped Si body, have high mobility, reduced low-frequency noise and short-channel effects, increased the threshold voltage and avoid effects due to depletion of poly-Si gates. The Ion/Ioff ratio of the device can be larger than 10^6, a key parameter for device operation.

Introduction

The continuously scaled gate length in conventional planar CMOS transistor is increasingly difficult in maintaining high drive currents with low off-current leakage as well as stability of threshold voltage. The short-channel effect is known degrading device performance and sets a limit for further scaling down conventional planar CMOS devices. One key development for further scaled down CMOS transistors is a stronger control of the channel conductance by the gate electrode instead if influence of the fringe field from the drain electrode. For SOI devices, from partial depletion mode to full depletion mode, this is achieved by reducing the silicon body thickness. Double gate transistor[1] and tri-gate transistor[2] are alternative device structures with stronger gate control over channel area (than planar CMOS) and lead to excellent scalability into short channel beyond the 32nm node. Gate-All-Around FinFET is one of the most promising structures to extend the scaling of the CMOS devices as it provides the best channel electrostatic control, which improves further with the shrinking of channel thickness.[3-5] Recently, there are numerous interests in SOI MOSFETs working in accumulation mode.[6-8] Similar to thin-body SOI MOSFETs in volume inversion mode, thin-body SOI can also operate in accumulation mode. The current in accumulation is flowing across the whole Si body, which is beneficial to increase the mobility, reduce the low-frequency noise and the short-channel effects as compared to FD SOI MOSFETs, increase the threshold voltage and avoid effects due to depletion of poly-Si gates. In this work, we propose a novel GAAC FinFET device with operation in accumulation mode, where the analytically calculated I-V characteristics agreed well with 3D TCAD simulation.

Device Architecture

Figure 1 illustrates a simplified perspective and cross-sectional view of a GAAC FinFET device structure. The source, drain and the channel regions are doped with the

Ex. 1016, Page 111

same type of dopant. Thus, there is no pn junction along with the channel length and reduce the leakage current. The device cross-sectional views along and perpendicular to the channel directions are shown in Figure 2. The ultimate GAAC FinFET device implements a physical oxide with a large bandgap to isolate the gate from the conducting channel area. By applying gate voltage to accumulate or deplete majority carriers in the channel, we can modulate the channel conductance for controlling the channel current as a switch between the source and drain.

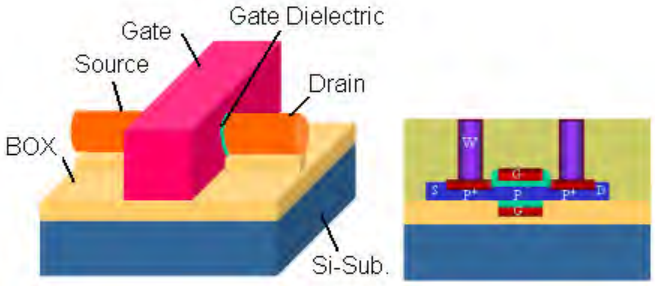


Figure 1 Simplified perspective (left) and cross-sectional (right) view of a GAAC FinFET device architecture direction

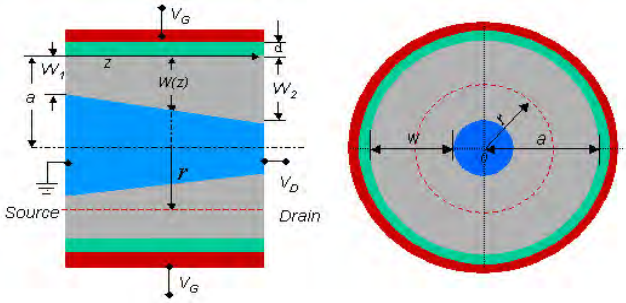


Figure 2 GAAC FinFET cross-sectional views: (a) along the channel direction; (b) perpendicular to the channel

Device Analysis

According to the symmetry aspect of the cylindrical device structure, we choose the cylindrical coordinate as described in Figure 2b for analysis. The surface potential and electric field distribution can be derived by solving the Poisson's equation in the silicon pillar. The electric field at a radial of r can be expressed by,

$$E(r) = -\frac{eN_A}{2\varepsilon_{si}}[r - \frac{1}{r}(a-w)^2], \ (0<r\leq a) \qquad [1]$$

where a is the radius of the cylinder channel; N_A is the density of acceptor impurity atoms; e is electronic charge; ε_{si} is the relative dielectric constant of silicon; w is the depletion width of the semiconductor. The potential drop in the semiconductor ψ(r) is given by,

$$\psi(r) = V(z) - \frac{eN_A}{4\varepsilon_{si}}[r^2 - (a-w)^2 + 2(a-w)^2 \ln(\frac{a-w}{r})] \qquad [2]$$

thus, the potential drop in the semiconductor surface ψ_s is given by,

$$\psi_s = \psi(a) = V(z) - \frac{eN_A}{4\varepsilon_{si}}\{a^2 + (a-w)^2[2\ln(1-\frac{w}{a}) - 1]\} \qquad [3]$$

where V(z) is the drain-to-source voltage drop along the channel. The current-voltage (I-V) characteristics of the GAAC FinFET can be analyzed as below. The differential

Ex. 1016, Page 112

resistance (dR) of the channel with the differential length (dz) at a point z in the channel is [9]

$$dR=\rho dz/A(z) \quad [4]$$

where ρ is the resistivity and $A(z)$ is the current flow through the cross-sectional area. If we neglect the minority carrier electrons in the p channel, the channel resistivity is

$$\rho=1/(e\mu_p N_A) \quad [5]$$

The cross-sectional area is given by

$$A(z)=\pi[a-w(z)]^2 \quad [6]$$

μ_p is the hole mobility, $w(z)$ is the depletion width at point z. Thus, the differential resistance of the channel can be expressed by

$$dR=dz/[\pi e\mu_p N_A (a-w)^2] \quad [7]$$

The differential voltage across a differential length dz can be expressed as

$$dV(z)=I_D dz/[\pi e\mu pNA (a-w)^2] \quad [8]$$

or

$$I_D dz =\pi e\mu_p N_A (a-w)^2\, dV(z) \quad [9]$$

where the drain current I_D is constant through the channel. The relationship between the potential in the channel due to the drain-to-source voltage $V(z)$ and the depletion width $w(z)$ is given by

$$V_g = V_{fb}+V_{ox}+\psi_s = V_{fb}-\frac{eN_A}{2\varepsilon_{ox}}w(2a-w)\ln(1+\frac{d}{a})+V(z)-\frac{eN_A}{4\varepsilon_{si}}\{a^2+(a-w)^2[2\ln(1-\frac{w}{a})-1]\} \quad [10]$$

where V_{ox} and V_{fb} are the potential drop in the gate oxide and the flat band voltage respectively; ε_{ox} is the dielectric constant of silicon oxide; d is the gate dielectric thickness; Equation [9] becomes,

$$I_D dz = \pi e\mu_p N_A (a-w)^2[\frac{eN_A}{\varepsilon_{ox}}(a-w)\ln(1+\frac{d}{a})-\frac{eN_A}{\varepsilon_{si}}(a-w)\ln(1-\frac{w}{a})]dw \quad [11]$$

Assuming the current and the mobility are constant through the channel, I_D can be derived by integrating equation [11] along the channel length (L) as below,

$$I_D = \frac{1}{L}\int_{W_1}^{W_2}\pi e\mu_p N_A (a-w)^2[\frac{eN_A}{\varepsilon_{ox}}(a-w)\ln(1+\frac{d}{a})-\frac{eN_A}{\varepsilon_{si}}(a-w)\ln(1-\frac{w}{a})]dw \quad [12]$$

Finally, it yields,

$$\begin{aligned} I_D = {} & \frac{\pi(eN_A)^2\mu_p}{4L}\{(a-w_2)^4\{\frac{1}{\varepsilon_{si}}[\ln(1-\frac{w_2}{a})-\frac{1}{4}]-\frac{1}{\varepsilon_{ox}}\ln(1+\frac{d}{a})\} \\ & -(a-w_1)^4\{\frac{1}{\varepsilon_{si}}[\ln(1-\frac{w_1}{a})-\frac{1}{4}]-\frac{1}{\varepsilon_{ox}}\ln(1+\frac{d}{a})\}\} \end{aligned} \quad [13]$$

where w_1 and w_2 represent the depletion width at the source and drain side and can be solved by the equation [14] and [15] respectively;

$$V_g = V_{fb}-\frac{eN_A}{2\varepsilon_{ox}}w_1(2a-w_1)\ln(1+\frac{d}{a})-\frac{eN_A}{4\varepsilon_{si}}\{a^2+(a-w_1)^2[2\ln(1-\frac{w_1}{a})-1]\} \quad [14]$$

$$V_g = V_{fb}+V_D-\frac{eN_A}{2\varepsilon_{ox}}w_2(2a-w_2)\ln(1+\frac{d}{a})-\frac{eN_A}{4\varepsilon_{si}}\{a^2+(a-w_2)^2[2\ln(1-\frac{w_2}{a})-1]\} \quad [15]$$

The drain current becomes saturated when the drain side is pinched-off, i.e. the depletion width at the drain side equals to the radius of the cylindrical channel. The saturation drain current is independent of the drain-to-source voltage in equation [16].

$$V_D = V_{Dsat} = \frac{eN_A a^2}{2\varepsilon_{ox}}\ln(1+\frac{d}{a})+\frac{eN_A a^2}{4\varepsilon_{si}}+V_G-V_{fb} \quad [16]$$

Figure 3 shows the calculated analytical curves of I_D versus V_D of a p-channel GAAC FinFET. The ideal current-voltage characteristics of I_D versus V_G on a p-channel GAAC FinFET is shown in Figure 4.

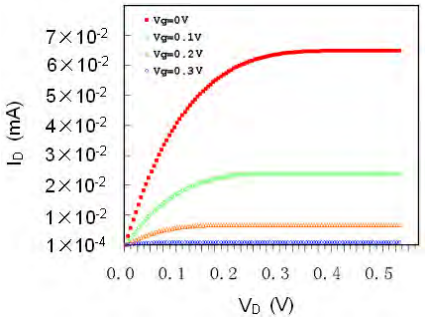


Figure 3 Calculated curves of I_D versus V_D of a p-channel GAAC FinFET with a=5nm, L=10nm, d=1nm, N_A=10^{19}cm^{-3}

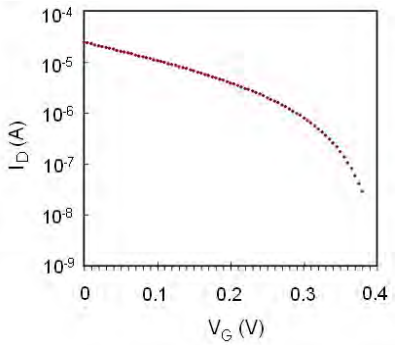


Figure 4 Calculated curves of I_D versus V_G of a p-channel GAAC FinFET with a=5nm, L=10nm, d=1nm, N_A=10^{19}cm^{-3}, V_D=-50mV

3D TCAD Simulation

As technologies advancing toward more complex, the semiconductor industry relies increasingly more on TCAD simulation for fast pace and lower cost in research and development on novel devices. Jimenez et al have proposed the analytical model and simulation on cylindrical Gate-All-Around MOSFETs focusing on inversion mode.[10-12] In this paper, SYNOPSYS 3D FLOOPS-DEVISE and DESSIS are used in our device simulation working in accumulation mode. A floating silicon body is applied to act as the device channel, with gate dielectrics and metal gate wrapped around. Doping dependence and surface roughness mobility degradation models are switched on in our simulation. Figure 5 shows the view of the simulated 3D GAAC FinFET structure with simulation grid.

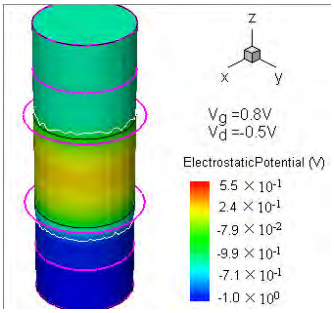


Figure 5 View of the simulated 3D GAAC FinFET structure on electrostatic potential distribution

Ex. 1016, Page 114

Figure 6 and Figure 7 show the simulated cross-sectional view on electrostatic potential and hole density distribution on a p-channel GAAC FinFET with Y-cut respectively. The cross-sectional view on simulated hole density for fully encapsulated GAAC FinFET at Z-cut at the middle of the channel position is given in Figure 8.

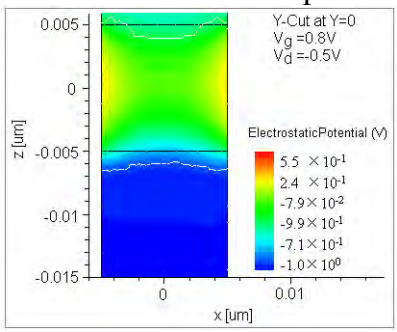


Figure 6 Cross-sectional view on simulated electrostatic potential for fully encapsulated GAAC FinFET at Y-cut with y=0; V_D and V_G are biased at –0.5V and 0.8V respectively; a=5nm, L=10nm, d=1nm, N_A=10^{19}cm^{-3}

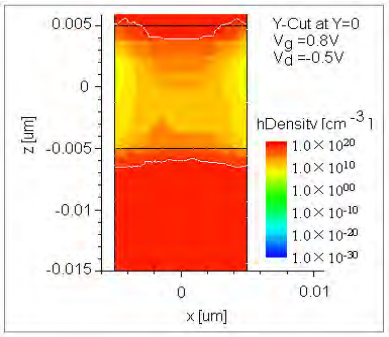


Figure 7 Cross-sectional view on simulated hole density for fully encapsulated GAAC FinFET at Y-cut with y=0; V_D and V_G are biased at –0.5V and 0.8V respectively; a=5nm, L=10nm, d=1nm, N_A=10^{19}cm^{-3}

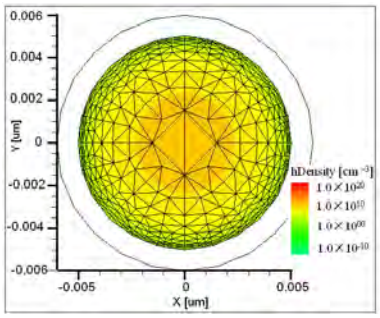


Figure 8 Cross-sectional view on simulated hole density for fully encapsulated GAAC FinFET at Z-cut with Z=0 (middle of the channel position) when V_D and V_G are biased at –0.05V and 0.8V respectively; a=5nm, L=10nm, d=1nm, N_A=10^{19}cm^{-3}

Figure 9 shows the simulated plot of I_D versus V_D on a p-channel GAAC FinFET. It agrees with our theoretical calculation results shown in Figure 3. In our simulation, the hydrodynamic and quantum potential have been implemented as physical models. Doping dependent, lattice scattering, surface scattering, carrier-carrier scattering, high field saturation have been taken into account in our mobility model. Shockley–Read–Hall recombination, Avalanche generation, Band to Band Tunneling (BTBT), Shockley-Read-Hall (SRH) recombination, Auger recombination are implemented in the Generation-Recombination model. The simulated current is slightly lower than that of the theoretical

calculation as it's more close to the real device operation. Further more, the impact of crystal orientation in cylindrical geometry, quantum effects and ballistic effects, etc are not included in this discussion.

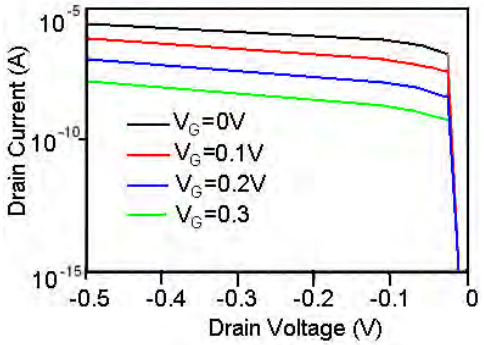


Figure 9 Simulated plot of I_D versus V_D on a p-channel GAAC FinFET

Conclusion

The GAAC FinFET device ideally provides the best gate electric field control as it has a virtually "infinite" number of gates with all the gates in close vicinity of the channel with enhanced electrostatic control from the gate electrode over the charge carriers in the channel. GAAC FinFETs operating in accumulation mode in which the current in accumulation is flowing across the whole cylinder-shaped Si body, have high mobility, reduced low-frequency noise and short-channel effects, increased the threshold voltage and avoid effects due to depletion of poly-Si gates. The Ion/Ioff ratio of the device can be larger than 10^6, which is a key parameter for device operation. The GAAC device can particularly further improve the device performance and the scalability over conventional multi-gate FinFETs.

References

1. Yu B, Chang L L, Ahmed S, et al., IEDM pp.251 (2002)
2. Doyle B, Boyanov B, Datta S, et al., Symposium on VLSI Technology Digest of Technical Paper, 10A-2 (2003)
3. Kranti A, Rashmi, Haldar S, Gupta R S, Electronics Letters, Vol. 37(12), pp.533-534 (2001)
4. Sang H O, Monroe D, Hergenrother J M, IEEE EDL Vol. 21(9), pp.445-447 (2000)
5. Krantia A, Haldarb S, Gupta R S, Microelectronics Journal Vol. 32(4), pp.305-313 (2001)
6. Park J W, Xiong W, Colinge J P, Proceedings of IEEE International SOI Conference, pp.65-67 (2003)
7. Afzalian A, Lederer D, Lee C W, et al., Proc. 4th EuroSOI Workshop, pp.47-48 (2008)
8. Afzalian A, Lederer D, Lee C W, et al., IEEE 2008 Silicon Nanoelectronics Workshop, Honolulu, pp.1-6 (2008)
9. Donald A. Neamen, "Semiconductor Physics and Device: Basic Principles" (Third Edition), McGraw-Hill Companies, Inc., pp.582-584 (2003)
10. Jimenez D, Iniguez B, Sune J, et al., IEEE EDL Vol. 25(8), pp.571-573 (2004)
11. Hamid A E, Iniguez B, Guitart J R, et al., IEEE ED Vol. 54(3), pp.572-579 (2007)
12. Moldovan O, Iniguez B, Jimenez D, et al., IEEE ED Vol. 54(1), pp.162-165 (2007)

Ex. 1016, Page 116

ECS Transactions, 18 (1) 89-92 (2009)
10.1149/1.3096433 © The Electrochemical Society

Bit Cost Scalable (BiCS) Flash Technology for Future Ultra High Density Storage Devices

Akihiro Nitayama and Hideaki Aochi

Center for Semiconductor Research & Development,
TOSHIBA Corporation, Semiconductor Company
8 Shinsugita-Cho, Isogu-Ku, Yokohama 235-8522, Japan
Phone: +81-45-776-5600, E-mail: akihiro.nitayama@toshiba.co.jp

We have proposed Bit-Cost Scalable (BiCS) technology which realizes a 3D multi-stacked memory array with a few critical lithography steps regardless of the number of stacked layers to keep a drastically continuous reduction of bit cost. The cell array concept, fabrication process, and key features are presented.

Introduction

Three dimensional memories are gathering increasing attention as future ultra-high density memory technologies to keep the trend of increasing bit density and reducing bit cost. Several candidates of the technologies for the future three dimensional non-volatile memory have been proposed, as shown in Fig.1, such as cross point cell with resistive switching memory and diode [1], simply stacked two dimensional array structures [2,3], Bit-Cost Scalable (BiCS) flash technology [4,5] and VRAT with PIPE [6]. The disadvantage of the cross point cell and simply stacked two dimensional arrays is that the process cost to build one layer of memory array is relatively high, because each layer needs several lithography steps and the other processes. This is the reason why the efficiency of cost reduction by stacking layers of memory array is not so beneficial. On the other hand, the last two technologies listed above realize a multi-stacked memory array with only a few number of critical lithography steps regardless of the number of stacked layers, and make continuous bit cost reduction possible. In this paper, the concept and the other technology issues of BiCS flash are reviewed.

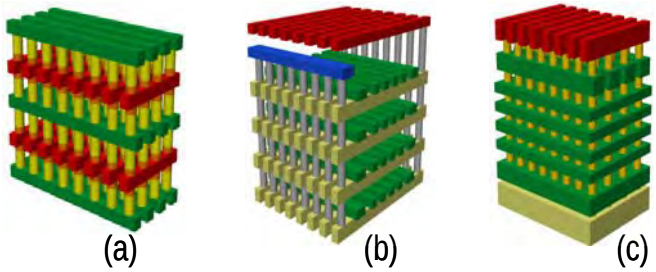


Fig.1. Examples of 3-D memory array (a) Cross point cell (b) Simply stacked planar cell and (c) BiCS flash.

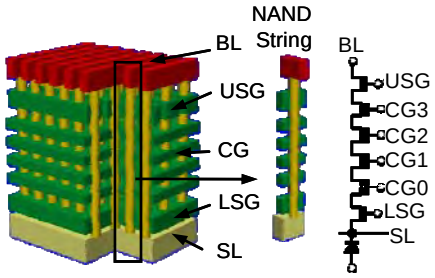


Fig.2. Concept of BiCS Flash

Concept of BiCS flash

Figure 2 shows the concept of BiCS flash memory. Whole stack of electrode plates are punched through and plugged with SONOS-type memory films and poly-silicon pillar at one time to form a series of vertical FETs which act as a NAND string of SONOS-type memories [4]. Several rows of NAND string share control gates and the lower select gate.

Ex. 1016, Page 117

Upper select gate is formed as line and space to take a role of row select line. Due to the punch and plug process, the sequence of FET formation is completely opposite to the conventional FET, i.e., gate dielectric films are deposited on the sidewall of poly-Si gate hole followed by deposition of channel poly-Si body as shown in Fig.3. The bottom of memory string is connected to the common source diffusion formed on the Si substrate. Figure 4 shows the birds-eye view and the cross section SEMs of fabricated test chip of BiCS flash with 4 layers of control gates.

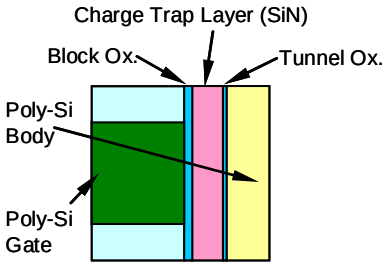


Fig.3. "Gate first" process for FET fabrication. Gate dielectric films are deposited on the Gate followed by channel Poly Si deposition.

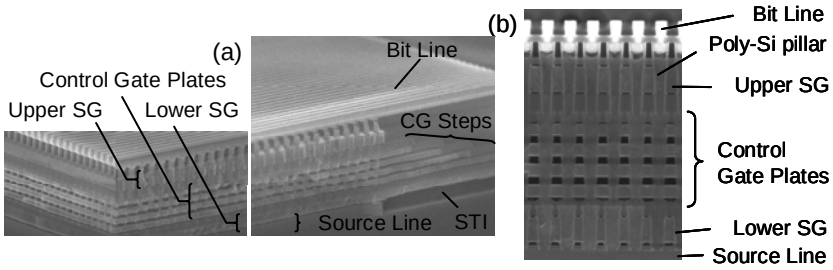


Fig.4. Birds-eye view (a) and cross section (b) SEM picture of fabricated BiCS flash in test chip.

Macaroni FET and SONOS-type memory films

As we reported, one of the most serious concern is that the channel of the vertical FET is poly-Si. To improve the FET characteristics, "Macaroni" type FET is proposed [5]. Figure 5 shows the schematic of "Macaroni" FET. By reducing the thickness of channel poly-Si less than the width of the depletion region induced by gate bias, the distribution of Vth is dramatically improved as shown in Fig.6. Figure 7 shows that the sub-threshold slope becomes steeper and the on-current increases with thinning Si thickness of "Macaroni FET".

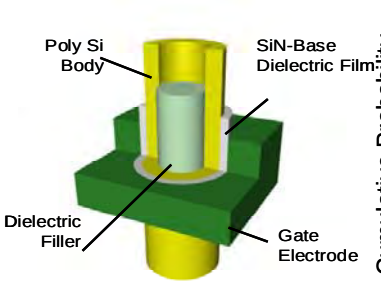


Fig.5. Schematic illustration of "Macaroni" vertical FET.

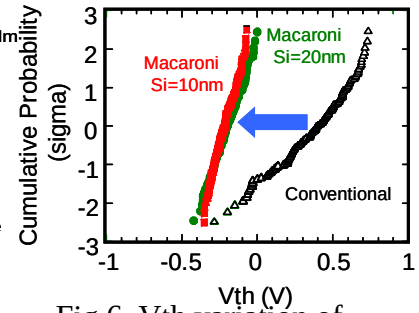


Fig.6. Vth variation of "Macaroni" Vertical FET. Thinner Si shows better results.

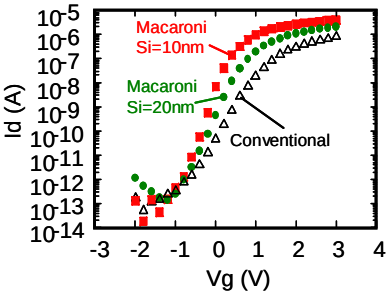


Fig.7.Id-Vg of "Macaroni" Vertical FET. Thinner Si shows better results.

We reported that a SiN-base gate dielectric film is better to get good compatibility with BiCS process integration [5]. The other important point for SONOS-type memory films is that the top down view of the string is a circle as shown in Fig.8. In this case, the radius of curvature of the tunnel oxide is less than that of the block oxide, that means the electric field on tunnel oxide is much larger than that of block oxide. As a result, BiCS does not need high-k material as block oxide even for the case that both of program and

Ex. 1016, Page 118

erase operation is done by FN current. The results by using the advanced SONOS process with spacer method is shown in Fig.8. The retention result is also excellent as shown in Fig.9.

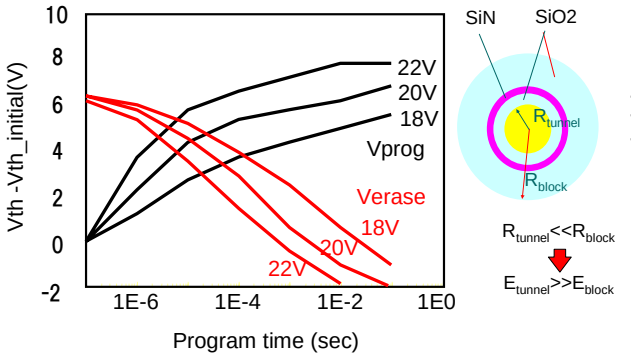


Fig.8. Program/Erase characteristics of BiCS with oxide block film. Radius of curvature difference between tunnel ox. and block ox. enables erase operation by FN current.

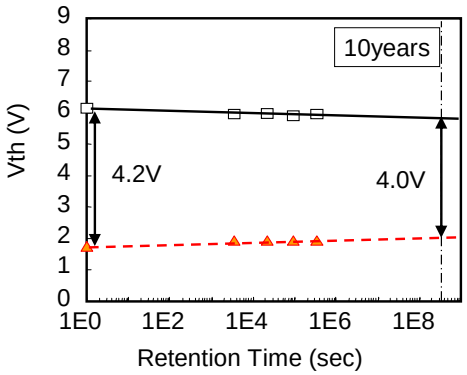


Fig.9. Data retention characteristics of BiCS with oxide block film.

Operation and Disturb for BiCS flash

A single memory cell is accessed at the intersection of a control gate and a NAND string, which is selected by a bit line and an upper select gate. The erase operation is done with holes generated by GIDL at the junction edge of upper and lower select gates as shown in Fig.10. Figure 11 shows that a higher junction concentration generates more GIDL, holes and ΔVth. The result of program and erase for each single cell on a string in the fabricated test chip is shown in Fig.12.

In the case of BiCS flash, there are many unselected strings which are stressed by Vprog, Vpass or Vread at the same time as the selected string, because some rows of strings are sharing control gates. However, high enough voltage boost happens due to no coupling between pillar and substrate when both upper and lower select gates isolate the pillar from bit line and source line well, so that the disturbs on the unselected string are not of serious concern[7].

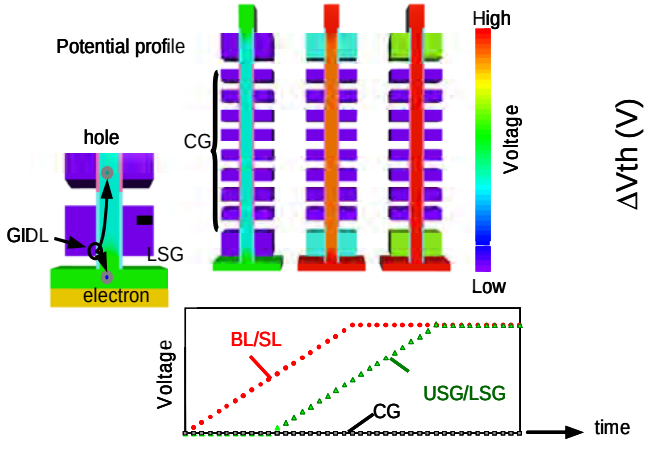


Fig.10. Schematic illustration of erase operation of BiCS Flash. Hole generated by GIDL at the USG & LSG edge raises the potential of pillar up.

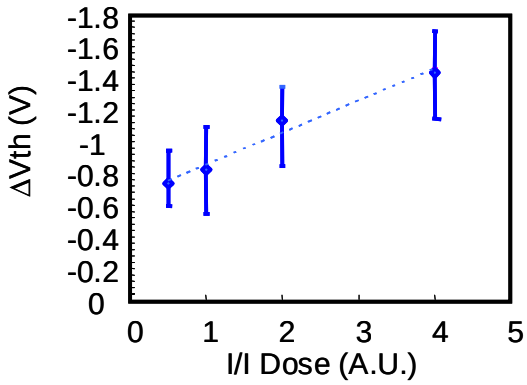


Fig.11. Vth shift by erase bias as a function of source/drain I/I dose. A higher junction concentration generates more GIDL, holes and larger Vth shift.

Conclusions

BiCS flash is one of the most cost effective memories in the currently proposed three dimensional non-volatile memories. It does not require new material for memory cell and there is no serious concern for disturbs. It is one of the promising candidates for the future ultra high density storage devices.

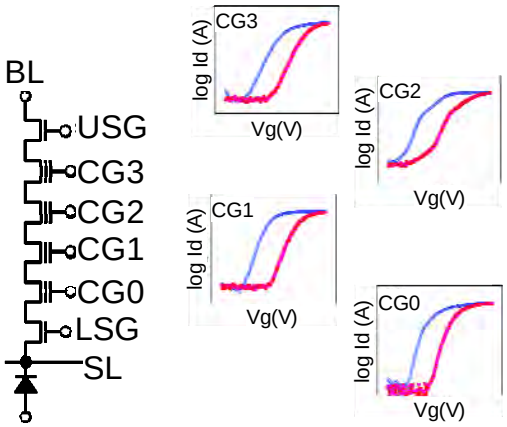


Fig.12 Results of program and erase for each single cell on a string in the fabricated test chip.

Acknowledgments

The authors would like to thank our team, including R.Katsumata, Y.Fukuzumi, M.Kito, M.Kido et al. for their great efforts in device fabrication, characterization, and fruitful discussions.

References

1. I.G.Baek et al. *IEDM Tech. Dig.*, 769-772 (2005).
2. Soon -Moon Jung et al. *IEDM Tech. Dig.*, pp 37-40 (2006).
3. Erh-Kun Lai et al. *IEDM Tech. Dig.*, pp 41-44 (2006).
4. H.Tanaka et al. *VLSI Symp. Tech Dig.* pp 14-15 (2007).
5. Y.Fukuzumi et al. *IEDM Tech Dig.*, pp 449-452 (2007).
6. J.Kim et al. *VLSI Symp. Tech Dig.*, pp122-123 (2007).
7. Y.Komori et al. *IEDM Tech Dig.*, pp 851-854 (2008).

Ex. 1016, Page 120

ECS Transactions, 18 (1) 93-98 (2009)
10.1149/1.3096434 © The Electrochemical Society

Silicon Nanocrystals Synthesized by Electron-beam Co-evaporation Method and Their Application for Nonvolatile Memory

Chen Chen [a], Rui Jia [a], Ming Liu [a], Wei Long Li[a], Chen Xin Zhu [a], Hao Feng Li [a], Pei Wen Zhang [a], Chang Qing Xie [a], Qin Wang [a], and Tian Chun Ye [a]
[a] Microfabrication and nanotechnology Lab, Institute of Microelectronics, Chinese Academy of Sciences, Beijing 10029, China

The Silicon nanocrystals designed for nonvolatile memory applications are for the first time prepared by electron beam co-evaporation of Si and SiO_2 hybrid. Transmission electron microscopy observation and Raman scattering certify the formation of Si NCs. Metal–oxide–semiconductor capacitor structures with Si NCs embedded in the gate oxide are fabricated to characterize the memory behavior. High-frequency capacitance-voltage and capacitance-time measurements demonstrate the memory behavior of the structure resulting from the charging/discharging behaviors of Si NCs. It is found that the flat-band voltage can be changed by adjusting the concentration of Si and SiO_2 in the hybrids. The Si nanocrystals memory device has yielded good retention characteristics with small charge loss.

I. INTRODUCTION

In order to overcome scaling down limitations of conventional nonvolatile memory (NVM), silicon nanocrystals (Si NCs) floating-gate memories have recently attracted extensive attentions (1-2), due to their promising characteristics, for example, compatibility with standard silicon technology, low P/E voltage with high programming speed, and long retention times (3). Practically, the formation of a Si NC layer as floating gate dielectric is the critical process for device performance. Up to now, several technologies have been employed to synthesize the Si NCs, including chemical vapor deposition (CVD) (4-6), Si ion implantation into SiO_2 film (7-8), and co-sputtering (9). However, there is no report about the synthesization of Si NCs by electron-beam co-evaporation (EBCE) and their application to floating gate memory. It is known that electron beam evaporation, compared to other techniques, has the following advantages: it is fully compatible with standard silicon CMOS processing; it is an efficient way of evaporating a source material and allowing for thin film coatings on a substrate; besides that, it can evaporate most materials. Hence, it will be of great potential to synthesize Si NCs by co-evaporating designed Si and SiO_2 hybrids.

In this report, Si NCs, for the first time synthesized by EBCE of Si and SiO_2 mixture, were embedded in SiO_2 matrix of the Metal-Oxide-Semiconductor (MOS) capacitor devices to study their NVM characteristics. Besides that, the influence of the concentration in the Si & SiO_2 hybrids on the memory characteristics of MOS capacitors is also experimentally demonstrated and discussed.

II. EXPERIMENT

Fig. 1 (a) shows the schematic cross-sectional structure of the MOS capacitor discussed in this paper. Firstly, the P type silicon (100) substrates with resistance of

Ex. 1016, Page 121

$2-3\Omega\cdot cm$ were chemically cleaned by standard CMOS process. A 5 nm SiO_2 film was thermally grown on p-type Si (100) wafers in a dry oxygen ambient at 950 ℃. Then, three types of Si & SiO_2 hybrids, denoted as Sample A, B and C, with different Si mass ratios of 22 at.%, 32 at.%, and 58 at.% were deposited by EBCE at the speed of 0.1 nm·sec^{-1} to a thickness of 25 nm. After the deposition, samples were annealed in N_2 ambient at 1050 ℃ for 1.5 hours. Subsequently, a 10 nm silicon oxide layer as the control gate dielectric was deposited by Low Pressure Chemical Vapor Deposition at 500 ℃. The square-shaped Al top-electrodes were defined by standard photolithograph and the lift-off process. After that, the Al back-electrode was evaporated on the rear surface of the silicon substrate. Finally, the rapid thermal annealing process was carried out for the alloying. In order to do the comparison, we also simultaneously fabricated the MOS structure with only 25 nm SiO_2 matrix layer.

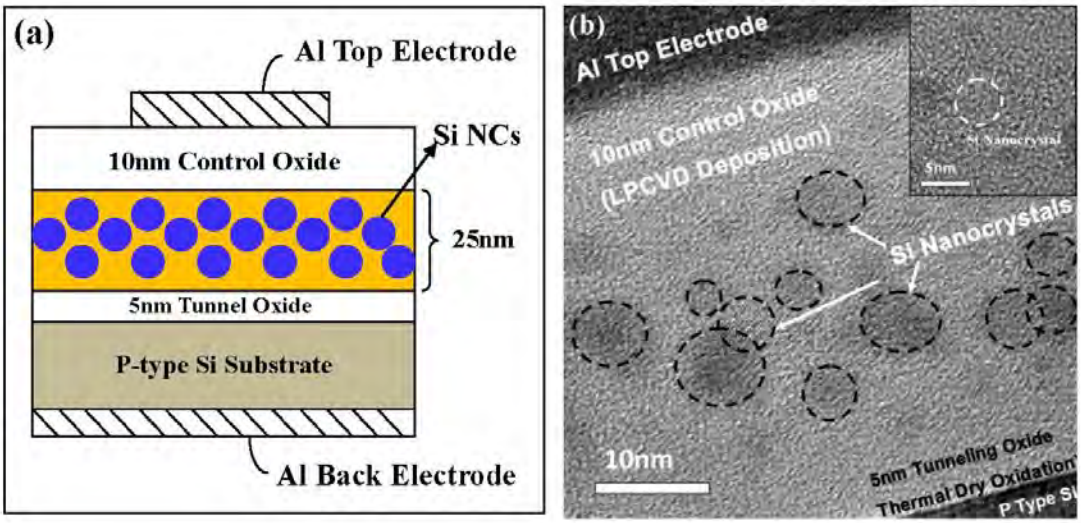


Fig.1 (a) Schematic view of MOS capacitor with Si NCs embedded in the gate oxide. (b) Cross-sectional TEM image: illustration of the MOS capacitor device structure with inset image of single Si nanocrystal in detail.

III. RESULTS AND DISCUSSION

The existence of Si NCs embedded in the SiO_2 matrix after annealing is clearly observed by cross-sectional transmission electron microscopy (TEM), as demonstrated in Fig. 1 (b). The high resolution TEM image of the single Si NC, shown in the inset, reveals distinct lattice fringe patterns, indicating the highly crystalline nature of the Si NCs. According to the TEM image, the nanocrystals density and average diameter can be determined as 3×10^{11} cm^{-2} and 4.5 nm, respectively, which best serves as the practical application of nanocrystals memory devices (10). From Fig. 1 (b), we can also see the tunnel oxide, control oxide and interfacial layers. The Si NCs are randomly distributed in the SiO_2 matrix with a thickness of 25 nm.

In order to further characterize the formation of Si NCs after annealing, Raman scattering measurement is performed, as shown in Fig. 2. The as-deposited sample shows the absence of a Raman signal. But for the sample annealed, a clear and sharp Raman peak situated at 517.7 cm^{-1}, the typical Raman spectrum of Si NCs, is observed. According to Zi et al. (11), the Raman shift of Si NCs can be predicted using the Bond

Ex. 1016, Page 122

Polarisability Modelling. In terms of this model, the Si NC diameter can be determined as about 4.4 nm, which is consistent with the TEM observation.

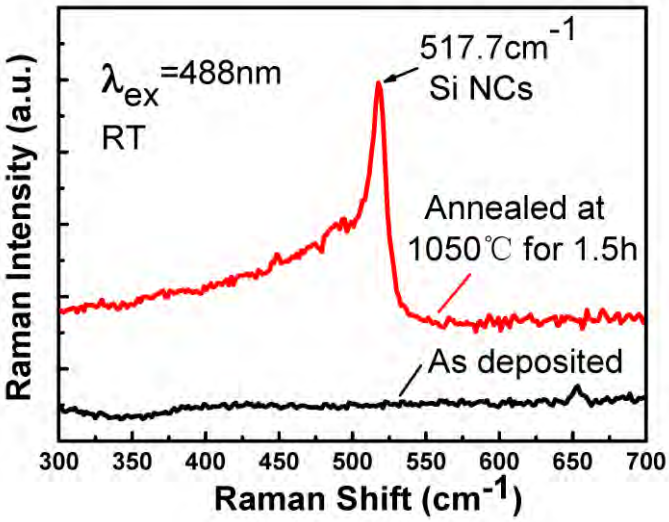


Fig.2 Raman Scattering spectra of the sample with Si NCs and the as-deposited sample, clear and sharp Raman peak from the Si NCs situated at 517.7 cm^{-1}.

Fig. 3 (a) shows the high-frequency (1 MHz, 300 K) C-V curves of MOS devices embedded with Si NCs under different sweeping voltages. As shown in inset plot of Fig. 3 (a), the C–V curve of the sample without Si NCs shows negligible loop hysteresis, indirectly proving that charging effect only occurs with Si NCs. But for the sample with Si NCs, a conspicuous counterclockwise hysteresis is observed. Here, it should be mentioned that the centre of the hysteresis curve is around −4 V, which, in ideal conditions, should be near 0 V. This shift may be induced by the excessive immobile positive charges existing in the oxide layer and the work function difference between the Al electrode and silicon (12). In Fig. 3 (a), the observed hysteresis loop is attributed to the charging and discharging process in the MOS capacitor structure by Fowler-Nordheim tunneling. The left four branches represent electron discharging while the right four branches represent electron charging. The charging or discharging of electrons leads to flat-band voltage shift (ΔV_{FB}), so called memory window. Thus, the memory window is dependent on the sweeping voltage range. According to Fig. 3 (a), the memory window is about 1.9 V when sweeping from -15 V to 13 V, 1.7 V when from -13 V to 10 V, and 1.5 V when from -10 V to 6 V, respectively. The shift in flat-band voltage can be used to evaluate the number of trapped charges theoretically by the following expression:

$$\Delta V_{FB} = \frac{Q_t}{C_{i0}} \quad (1)$$

Where C_{i0} (1.4×10^{-7} $F\cdot cm^{-2}$, according to our experimental result) is the area capacitance of the MOS structure, and Q_t is the area density of the trapped charges. For a memory window of 1.9 V obtained with a sweeping voltage from -15 V to 13 V, the Q_t can be estimated as 2.7×10^{-7} $C\cdot cm^{-2}$, therefore, giving the area electron density of 1.7×10^{12} cm^{-2}. Considering the density of NCs obtained by TEM image, it is expected that approximately six electrons were trapped in each nanocrystal.

To study the influence of the Si & SiO_2 concentration on the memory behavior, three types of MOS capacitor samples A, B and C with different Si mass ratios in hybrids are measured and compared, as shown in Fig. 3 (b). It can be seen that the Si mass ratio has

Ex. 1016, Page 123

directly influence on the memory window. With the increase of Si mass ratio, the memory window also increases, for instance, 0.3 V for sample A, 1.4 V for sample B and 1.6 V for sample C. The inset of Fig. 3 (b) exhibits that both the memory window and the Si NCs area density are directly proportional to the Si mass ratio. This may be due to the fact that Si NCs could be easily formed with higher Si concentration.

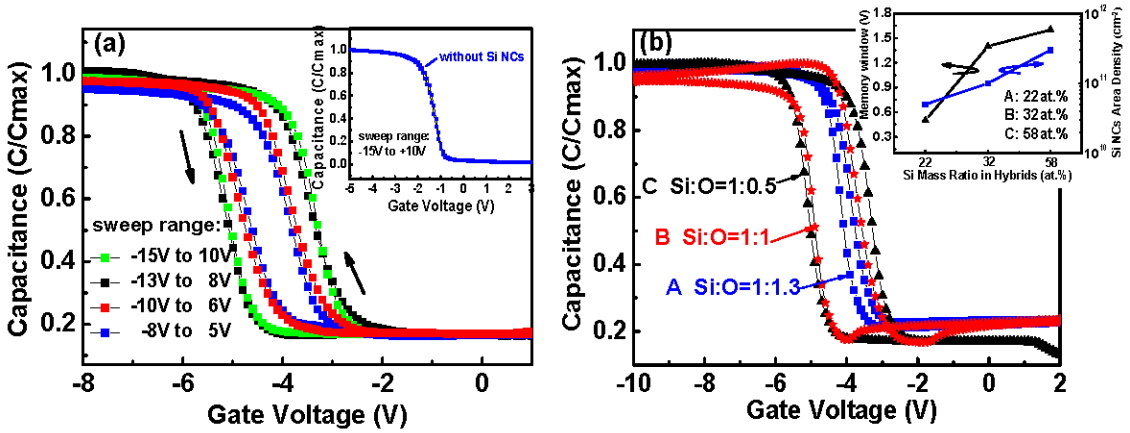


Fig.3 High frequency (1MHz) C-V measurement. (a) C-V curves of the sample C under different bi-direction sweeping voltages. The inset plot shows the C-V curves without Si NCs in SiO2 matrix, and (b) C-V hysteresis of Samples A, B and C with different densities of Si NCs under the same sweeping voltage range. The inset shows the tendency of memory window and Si NCs area density against Si mass ratio in hybrids.

To further analyze the retention characteristics, the charge retention time at room temperature is measured for the sample C under a fixed gate bias at flat-band voltage after programming. Fig. 4 displays the capacitance–time (C-T) curve for the MOS capacitor with Si NCs. The evolution shows the capacitance that is completely stable with the time before 10^4 s, indicating few charges has been leaked during that time. If the retention time is defined as the time when 5 % of the stored electrons have been leaked (13), the retention time is experimentally found to be exceeded up to 10^4 s, which is very promising for the possible NVM applications.

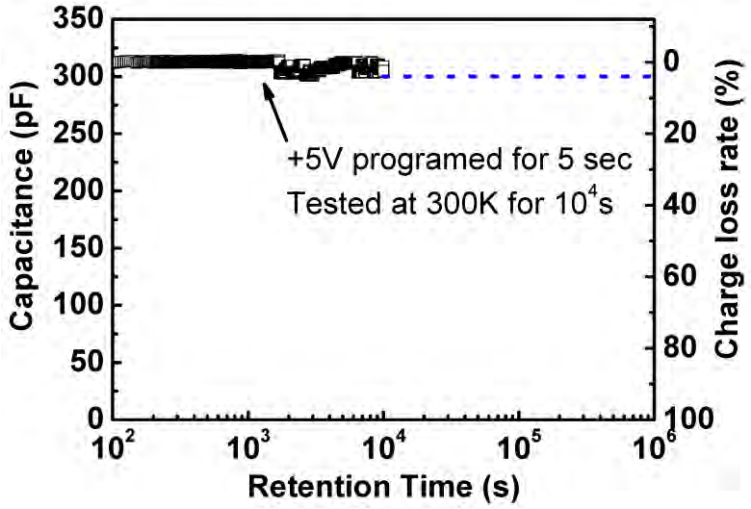


Fig. 4 Si NCs retention characteristics of fabricated memory MOS structure.

IV. CONCLUSION

The Si NCs embedded in SiO_2 matrix is successfully synthesized by electron beam co-evaporation of the Si & SiO_2 hybrid. The formation of Si NCs has been confirmed by TEM observation and Raman scattering measurement. The experiments show that by this way, high-density, uniform Si NCs can be obtained. The MOS capacitor devices with Si NCs area density of 3×10^{11} cm^{-2} and diameter of 4.5 nm clearly show a large memory window, which can be flexibly adjusted by changing the Si mass ratio in Si & SiO_2 hybrid. The measurement of MOS capacitor devices embedded with Si NCs shows long charge retention time of up to 10^4 s, which is promising for nonvolatile memory applications.

Acknowledgments

This work is subsidized by the 973 Projects 2006CB604904, 2006CB302706, by the Chinese NSF under Grant Nos.60706023, 60676001, 90401002, 90607022, by CAS YZ0635 and by CAS “Bai Ren Ji Hua”.

References

1. Sandip Tiwari, Farhan Rana, Kevin Chan, Hussein Hanafi, Wei Chan, and Doug Buchanan, in IEDM Tech. Dig., pp. 521-524 (1996).
2. J. De Blauwe, M. Ostraat, M. L. Green, G. Weber, T. Sorsch, A. Kerber, F. Klemens, R. Cirelli, E. Ferry, J. L. Grazul, F. Baumann, Y. Kim, W. Mansfield, J. Bude, J. T. C. Lee, S. J. Hillenius, R. C. Flagan, and H. A. Atwater, in IEDM Tech. Dig., pp. 29.7.1-29.7.4 (2000).
3. B. De Salvo, C. Gerardi, S. Lomhardo, T. Baron, L. Pemiola, D. Mariolle, P. Mur, A. Toffoli, M. Gely, M .N. Semena'l, S. Deleonibus, G. Ammendola, V. Ancarani, in IEDM Tech. Dig., pp. 26.1.1-26.1.4 (2003).
4. David Comedi, Othman H, Y. Zalloum, Jacek Wojcik, and Peter Mascher, IEEE Journal of selected topic in quantum electronics, VOL. 12, NO. 6, pp. 1561-1569 (2006).
5. T. Baron, A. Fernandes, J. F. Damlencourt, B. De. Salvo, F. Martin, F.Mazen, and S. Haukka, Appl. Phys. Lett., 82, 4151 (2003).
6. T. Z. Lu, M. Alexe, R. Scholz, V. Talalaev, R. J. Zhang, and M. Zacharias, Journal of Applied Physics, 100, 014310 (2006).
7. P. Dimitrakis, E. Kapetanakis, D. Tsoukalas, D. Skarlatos, C. Bonafos, G. B. Asssayag, A. Claverie, M. Perego, M. Fanciulli, V. Soncini, R. Sotgiu, A. Agarwal, M. Ameen, C. Sohl, and P. Normand, Solid State Electron., vol. 48, no. 9, pp. 1511–1517 (2004).
8. C. Y. Ng, T. P. Chen, L. Ding, and S. Fung, IEEE Electron Device Lett., vol.27, NO.4, pp. 231- 233 (2006).
9. S. Hayashi, and K. Yamamoto, Journal of Luminescence vol. 70, pp. 352-363 (1996).
10. Min She, and Tsu-Jae King, IEEE Transactions On Electron Devices, Vol. 50, No. 9, pp. 1934- 1940 (2003).
11. J. Zi, H. Buscher, C. Falter, W. Ludwig, K. Zhang, and X. Xie, Appl. Phys. Lett, 69, 200 (1996).

12. S. M. Sze, "Physics of Semiconductor Devices, Second Edition," New York: Wiley, pp. 390-402 (1998).
13. Min She, Ya-Chin King, Tsu-Jae King, and Chenming Hu, "Modeling and design study of nanocrystal memory devices," Device Research Conference, pp. 139-140 (2001).

ECS Transactions, 18 (1) 99-104 (2009)
10.1149/1.3096435 © The Electrochemical Society

Investigations on the Impact of the Parasitic Bottom Transistor in Gate-All-Around Silicon Nanowire SONOS Memory Cells Fabricated on Bulk Si Substrate

Yujie Ai, Ru Huang*, Yiqun Wang, Jing Zhuge, Dake Wu, Runsheng Wang, Poren Tang, Lijie zhang, Zhihua Hao, Yangyuan Wang

Institute of Microelectronics, Peking University, Beijing 100871, P. R. China

Gate-all-around (GAA) Si nanowire SONOS memory cells (SNWMs) have been fabricated on Si substrate using fully epi-free compatible CMOS technology. A parasitic bottom SONOS memory (PBM) was formed when the SNWM was fabricated on bulk Si substrate. The impact of the PBM on the performance of the SNWM is investigated in this paper. The PBM shows a slower program speed, a faster erase speed, and worse retention characteristics than the SNWM. Therefore, the PBM severely degrades the performance of the SNWM due to its slower program speed and worse retention characteristics, and should be carefully controlled for the SNWM based on bulk Si substrate.

Introduction

New device structures such as FinFET and Tri-gate have been successfully demonstrated suitable for the next generation flash memory devices due to their improved program/erase characteristics (1-5). Very recently, Gate-all-around(GAA) Si nanowire SONOS memories (SNWMs) have been reported, which exhibit superior fast program/erase speed (P/E) and better channel controllability over other new structure SONOS memories (6, 7). However, most of SNWMs utilize SOI wafers or selective epitaxial process, which increases the fabrication cost and complexity. Therefore, it is beneficial to fabricate SNWM on fully bulk Si substrate. However, the parasitic bottom SONOS memory cell (PBM) will be fabricated for the SNWM based on bulk Si substrate. It has been reported that parasitic bottom transistor can rarely impact the DC characteristics of the nanowire transistor based on bulk Si substrate due to its high threshold voltage (8, 9). However, the influence of the PBM on the performance of SNWMs is unclear. In this paper, the impact of the PBM on the P/E and retention characteristics of the memory cell is studied in details. In addition, some methods have been proposed to suppress the negative impact of PBMs.

Fabrication

Fig. 1 shows the schematic view of the GAA SNWM. The process flow of the SNWM is similar to the previously reported Si nanowire transistor (9). A $SiO_2/SiN/SiO_2$ stack was formed as the gate dielectric, as shown in Fig. 1(b) and (c). The SiN (5nm) trap layer and the blocking oxide (12nm) were deposited by LPCVD, and the tunneling oxide (8nm) was grown by dry thermal oxidation of the Si nanowire. As shown in Fig. 1(b) and (c), the PBM was formed simultaneously when the SNWM was fabricated on bulk Si substrate. Since a high dose B^+ substrate implantation has been performed to suppress the parasitic bottom transistor, much thicker and poorer tunneling oxide of the PBM was

Ex. 1016, Page 127

formed in the high doping region (10). Moreover, many traps may be introduced to the tunneling oxide of the PBM during these processes or during P/E cycles (11).

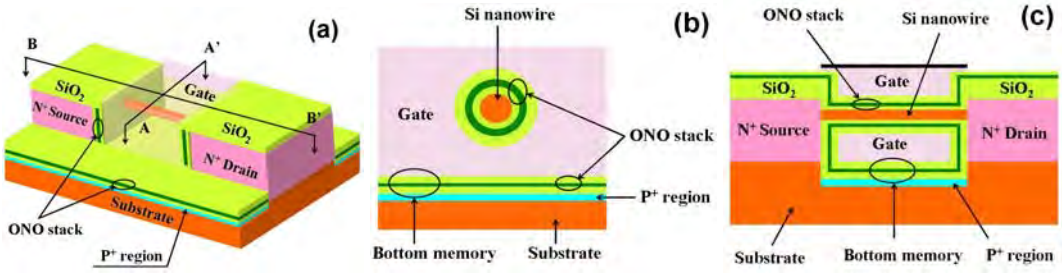


Figure 1. (a) Schematic view of the SNWM; (b) cross-sectional view along the A-A' direction; (c) cross-sectional view along the B-B' direction.

Results and Discussion

Figure 2 shows the measured I_d-V_g characteristics of the cell at P/E state. The FN tunneling was used for P/E by biasing the gate electrode positive or negative respectively, with source and drain grounded. A threshold voltage window of 2.5V is obtained with a 15V/1000μs program pulse and a -15V/10ms erase pulse.

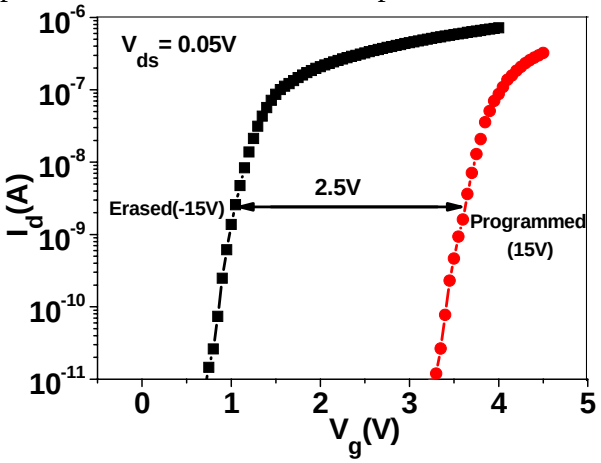


Figure 2. I_d-V_g characteristics of the cell at P/E state.

Figure 3 shows the program characteristics of the cell with different program time, which is obtained by programming the cell at 14V for 0μs, 100μs, 200μs and 1000μs, respectively. The drain current of the cell is composed of currents from both the SNWM and the PBM. In addition, the SNWM shows a much faster program speed than the PBM.

Ex. 1016, Page 128

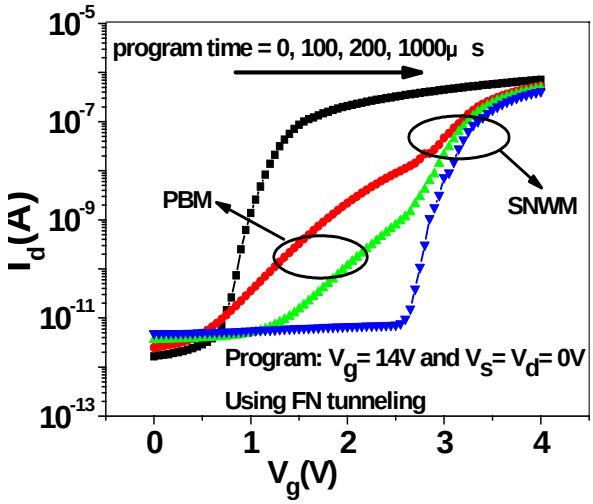


Figure 3. Program characteristics of the cell with different program time.

Figure 4 shows the program characteristics of the SNWM and PBM using FN tunneling with a program voltage of 14V. The PBM shows a much slower program speed than the SNWM. As a result, the PBM decreases the sense margin of the cell when the cell is read. In other words, the program time should be increased to improve sense margin of the cell due to the slow program speed of the PBM. Therefore, the PBM severely degrades the performance of the SNWM.

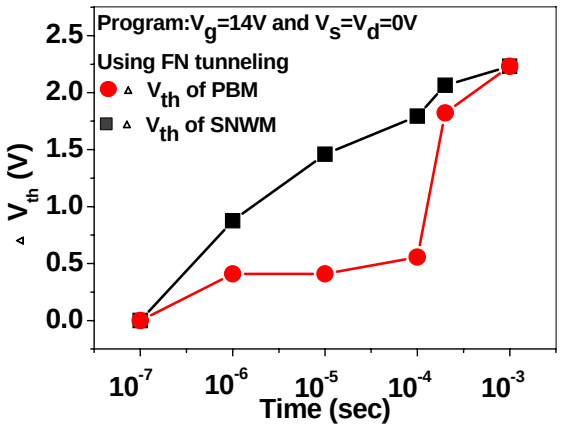


Figure 4. Program characteristics of the SNWM and PBM using FN tunneling.

Figure 5(a) and (b) show the schematic band diagrams of the SNWM and PBM when the cell is programmed, respectively. As shown in Fig. 5(a), the SNWM can exhibits a fast program speed using FN tunneling mechanism due to the effect of field enhancement (6, 7) and its thinner tunneling oxide. However, as shown in Fig. 5(b), the tunneling oxide of the PBM is 2~3 times (10) as thick as that of the SNWM. In addition, there is a high density of traps distributed in the tunneling oxide of the PBM. As a result, trap-assisted tunneling may become dominant for the PBM. When the PBM is programmed, electrons are captured by traps in both the nitride layer and the tunneling oxide layer. However, many electrons captured in these shallow traps of tunneling oxide will be detrapped by thermal excitation (12), and drift back to Si substrate again when transfer characteristics

of the cell for V_{th} extraction were measured. Therefore, this detrapping process severely degrades the program speed of PBMs.

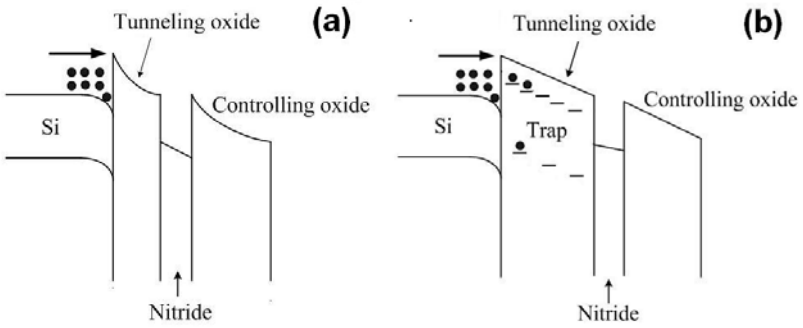


Figure 5. Schematic band diagrams of the SNWM (a) and PBM (b) when the cell is programmed.

Figure 6 (a) shows the erase speed characteristics of the SNWM and PBM using FN tunneling with an erase voltage of -15V. The PBM shows a faster erase speed than the SNWM. Fig. 6 (b) shows the schematic band diagram of PBM when the cell is erased. Trap-assisted tunneling mechanism is also dominant when the PBM was erased. The fast erase speed of the PBM may be resulted by the following two reasons. Firstly, the electrons in the shallow traps of tunneling oxide can be detrapped by thermal excitation and drift back to Si substrate again with the assistance of electric field (12). Secondly, the electrons trapped in the nitride layer can return to Si substrate again without tunneling through the whole tunneling oxide. Once the electrons are injected to the tunneling oxide and captured by these shallow traps, they can be detrapped by thermal excitation, and drift back to Si substrate again when transfer characteristics of the cell for V_{th} extraction were measured. Therefore, the PBMs exhibit a very fast erase speed due to the impact of shallow traps in its tunneling oxide.

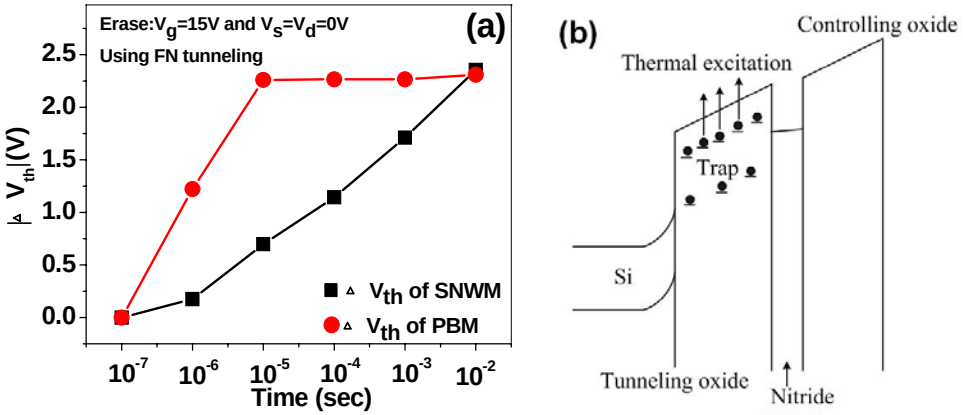


Figure 6. (a) Erase characteristics of the SNWM and PBM using FN tunneling; (b) Schematic band diagram of the PBM when the cell is erased.

Figure 7 shows charge retention characteristics of a programmed cell at room temperature. As shown in this figure, the PBM exhibits a much worse retention characteristics than the SNWM. The worse retention characteristics of the PBM may also be caused by the electron detrapping from the shallow traps in its tunneling oxide.

Moreover, the worse retention characteristics of the PBM will also result in a decrease of the sense margin of state ‘0’ and state ‘1’ when the cell is read.

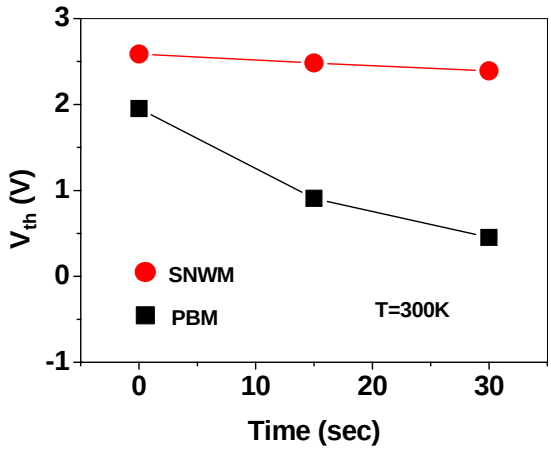


Figure 7. Charge retention characteristics of a programmed cell at room temperature.

As discussed above, the PBM severely degrades the performance of the SNWM due to its slow program speed and bad retention characteristics. Therefore, the PBM should be carefully controlled by optimizing the process flow of the SNWM. It is expected that the PBM can be successfully suppressed by using SDON (13) or SDOI (14) structure, which can be easily fabricated by additional implantation and annealing processes before the formation of Si nanowire.

Conclusion

GAA SNWMs have been fabricated on bulk Si substrate through epi-free CMOS-compatible process technology. A threshold voltage window of 2.5V can be achieved with a 15V/1000μs program pulse and a -15V/10ms erase pulse using FN tunneling mechanism. The impact of the PBM on the performance of the SNWM is investigated in details in this paper. The PBM shows a slower program speed, while exhibits a faster erase speed and worse retention characteristics due to the impact of traps in the thicker tunneling oxide. Therefore, the PBM severely degrades the performance of the SNWM, and should be specially designed.

Acknowledgement

This work was supported in part by the collaborative project between Samsung Electronics Co. and Peking University, the NSFC (60625403), 973 Projects (2006CB302701), and the NCET Program.

References

1. Jin-Woo Han and Seong-Wan Ryu, p. 200, Symp. VLSI Tech. Dig., (2008).
2. Jin-Woo Han and Seong-Wan Ryu, EDL., **29,** 632 (2008).
3. Jin-Woo Han and Seong-Wan Ryu, EDL., **29,** 781(2008).
4. Lombardo. S and Gerardi. C, p. 924, IEDM., (2007).
5. Tzu-Hsuan Hsu and Hang Ting Lue, EDL., **28,** 443(2007).

Ex. 1016, Page 131

6. Sung Dae Suk and Kyoung Hwan Yeo, p. 142, Symp. VLSI Tech. Dig., (2007).
7. J. Fu and K. D. Buddharaju, p. 79, IEDM Tech., (2007).
8. Sung Dae Suk and Sung-Young Lee, p. 717, IEDM Tech., (2005).
9. Y. Tian and R. huang, p. 895, IEDM Tech. Dig., (2007).
10. B E Deal and M Sklar, J. Electrochem. Soc., **112,** 430 (1965).
11. Neal Mielke and Hanmant Belgal, Transactions on Device and Materials Reliability, **4**, 335 (2004).
12. Ren-ichi Yamada and Yuki Mori, p. 200, 38th Annual International Reliability Physics Symposium, (2000).
13. Weihai Bu and Ru Huang, p. 269, ICSICT., (2004).
14. ZhikuanZhang and Shengdong Zhang, TED., **51,** 1095(2004).

ECS Transactions, 18 (1) 105-109 (2009)
10.1149/1.3096436 © The Electrochemical Society

Halo Implantation Optimization for 65nm Low Power PMOS Device Inverse Narrow Width Effect Improvement

Haohua Ye*, Jianhua Ju, Xuejie Shi, Bingwu Liu, Jay Ning, I. C. Chen
Semiconductor Manufacturing International Corporation,
18 Zhang Jiang Rd., Pudong New Area, Shanghai 201203, China
*Email: shirly_ye@smics.com

Device with shallow trench isolation (STI) has inverse narrow width effect (INWE), that is the threshold voltage of narrow width transistors is lower than that of large width transistors. The conventional methods of improving INWE focus on STI module optimization, such as active area corner rounding, step height and divot control. In this paper, a new approach by implant optimization is proposed to improve INWE. For low power PMOS device, phosphorus halo implant is often used for its lower diode leakage characteristics. This paper reports optimized dual halo implant approach with both arsenic and phosphorus implants to improve INWE. This approach gives larger Gm*Rout and lower digital product chip stand-by current without reliability degradation.

Introduction

STI-isolated device has inverse narrow width effect (INWE). That is threshold voltage of narrow width transistor is lower than that of wide width transistor, due to the two-dimensional (2-D) field–induced edge-fringing effect at the gate edge and hence formation of an inversion layer at the edges occurs at a lower voltage than the voltage required at the center. [1] This effect introduces higher off current in narrow devices, and causes higher standby power dissipation on chip. Therefore, INWE improvement is crucial for reducing chip standby power. The conventional methods for improving INWE are focused on STI edge profile optimization, such as corner rounding [2], reducing divot, and step height control.

In this paper, adjusting implantation species combination was used to improve INWE. For low power PMOS devices, phosphorus implantation is often used in order to have lower diode leakage instead of arsenic implantation. However, the phosphorous halo implantation of a 65nm low power PMOS process has shown a strong INWE. The process was then modified using dual halo implantation, which consists of arsenic implantation and phosphorus implantation. The results show that this new approach improves INWE effectively.

Experiments

PMOS surface channel device is fabricated using a 65nm logic process flow. A shallow trench isolation process defines device active area, followed by well formation by N-type ion implantation. Poly gate patterning follows the gate oxide insulator

Ex. 1016, Page 133

formation. A SiN based offset spacer is then fabricated, followed by source/drain extension and halo implantation. A poly gate spacer is formed and source/drain is defined by boron implantation and thermal activation. Finally, nickel silicide is fabricated for low contact resistance and subsequent contact and metal layers are formed, so electrode of the device is completed. Standard threshold voltage (SVT) device and High threshold voltage (HVT) device are fabricated in the same chip, the difference of these two type devices is the additional channel implantation on HVT device area.

In this paper, a single halo implantation approach using phosphorus implantation only (energy in a range of 10~30KeV, dosage in a range of 3E13~8E13 cm^{-2}) is used as control sample, two splits of the dual halo implantation approach including arsenic implantation (energy in between 30 and 60KeV, dosage in between 1E13 and 5E13 cm^{-2}) and phosphorus implantation (energy in between 10 and 30KeV, dosage in between 1E13 and 5E13 cm^{-2}) are fabricated. The split 1 uses high arsenic dosage combined with low phosphorus dosage. The split 2 adopts low arsenic dosage combined with high phosphorus dosage. Both of these two splits have comparable wide width device performance as compared to the control sample. Samples for these two approaches use same process conditions otherwise.

Transistor related electrical parameters are tested in an Agilent 4072, Vth_lin and Vth_sat were tested by a constant current method, sweeping Vgs and get Vth_lin=Vgs @ Ids=0.1uA*W/L &Vds=-0.1V and Vth_sat=Vgs @ Ids=0.1uA*W/L & Vds=-1.2V, respectively. Idsat is measured as Id @ Vgs=Vds=1.2V. Ioff is defined as Id @Vgs=Vb=0V & Vds=1.1*(-1.2)V. Gm* Rout data is the product of Gm and Rout, and Gm is defined as ΔId/ΔVg @ Vgs=Vds=Vt+0.24V, Rout is defined as ΔVd/ΔId @ Vgs=Vds=Vt+0.24V. PMOS Id~Vg curve data is collected in an Agilent 4156.

Results and Discussion

Arsenic and phosphorus dual halo implant approach shows a better INWE performance than a single halo implant (phosphorus only), as illustrated in Fig.1 and Fig.2. Higher arsenic dosage split has less Vth drop on Vth vs. W curve than that of low arsenic dosage split. For high arsenic dosage split,ΔVth_lin =Vth_lin (W/L=10/0.06um) - Vth_lin (W/L=0.12/0.06um) decreases by about 50 mV than phosphorus only control wafer for SVT device. The split reduces ΔVth_lin by 30mV for HVT device. INWE of Vth_sat has the same trend for the two splits and control sample. Generated from TCAD simulation, the As and P concentration under the channel edge with respect to the channel width is plotted in Fig. 3 for dual halo implant. Because of the P and As segregation into oxide from silicon, both dopant concentrations show a decreasing trend from the center to AA/STI interface at the channel width direction. One can also find that P concentration drops earlier than As does as close to the AA/STI interface at both 10nm and 20nm depth in the channel. Even though arsenic species has similar segregation coefficient as phosphorus, arsenic has much lower diffusivity than phosphorus has and hence much less arsenic ions diffuse to the silicon/STI oxide interface. Because less arsenic species out-diffuse into the STI oxide as compared to the phosphorus, dual halo implantation approach has higher Vth controllability in narrow width devices than the single P halo

Ex. 1016, Page 134

implant does. Shown in Fig. 4 is a comparison of total donor concentration along the channel width for both single and dual halo implant, where it can be clearly found that dual halo implant results in high average channel doping concentration hence high Vt.

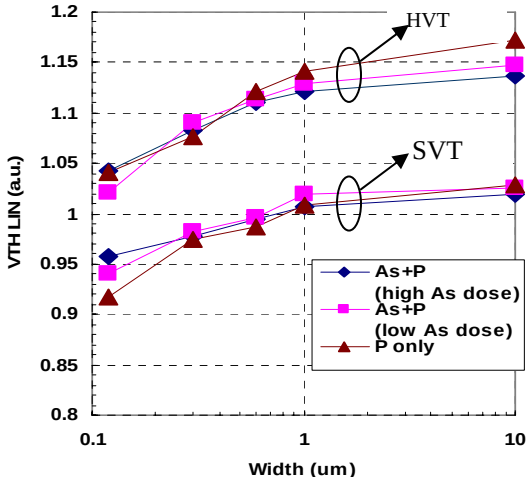


Fig. 1 PMOS device Vth_lin vs Width (Lg=0.06um)

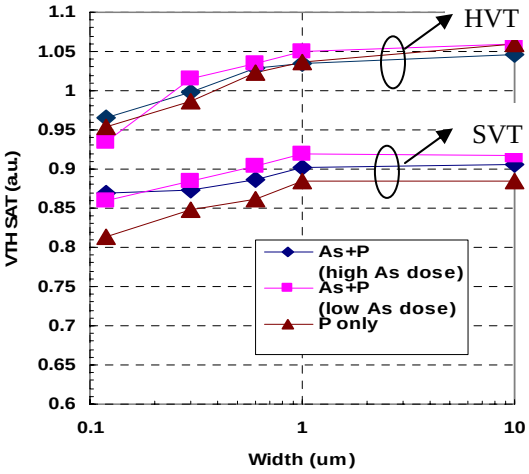


Fig. 2 PMOS device Vth_sat vs Width (Lg=0.06um)

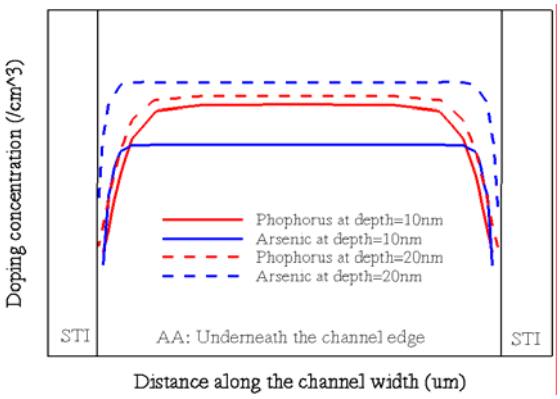


Fig. 3 TCAD simulated As and P doping concentration along the channel width under the channel edge using dual halo process

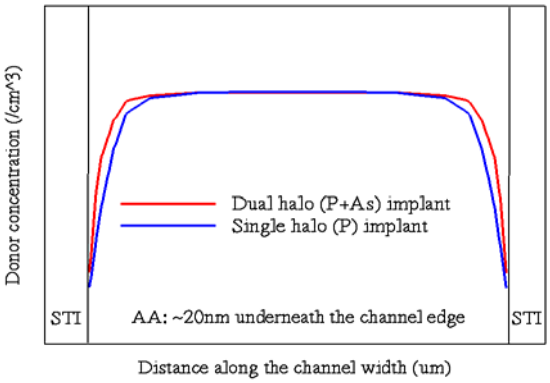


Fig. 4 TCAD simulated total donor concentration along the channel width under the channel edge using both dual and single halo implant

Fig. 5 shows that dual halo implantation reduces narrow width transistor Idsat both for SVT and HVT devices. Fig. 6 demonstrates that dual halo implantation can reduce narrow device Ioff for SVT devices, whereas increase narrow HVT devices Ioff. HVT device needs an additional channel implantation, and it's threshold voltage is about 150mV higher than that of SVT. The diode leakage for HVT is more important than that for SVT, since it has much lower sub- threshold leakage. Arsenic implantation introduces sharper dopant profile near the junction border and introduces more defects and hence subjects to more diode leakage. Therefore dual halo implantation would result in higher HVT device Ioff than phosphorus only, and higher arsenic dosage split has higher HVT device Ioff.

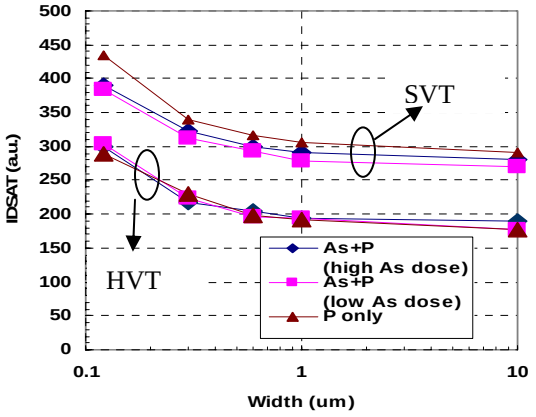


Fig. 5 PMOS device Idsat vs Width (Lg=0.06um)

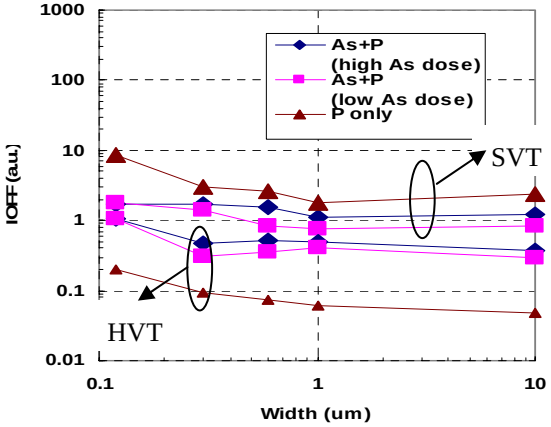


Fig. 6 PMOS device IOFF vs Width (Lg=0.06um)

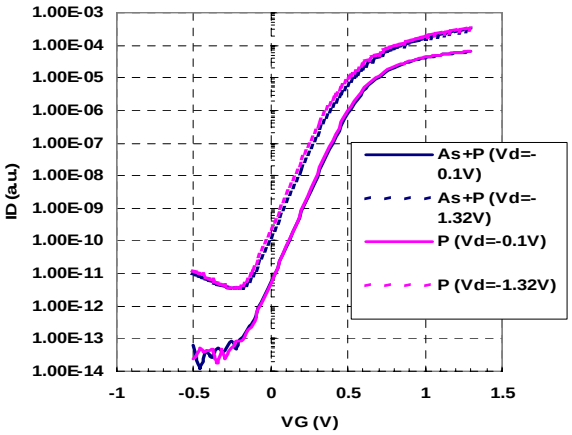


Fig. 7 PMOS ID vs VG curve (W/L=10/0.06um)

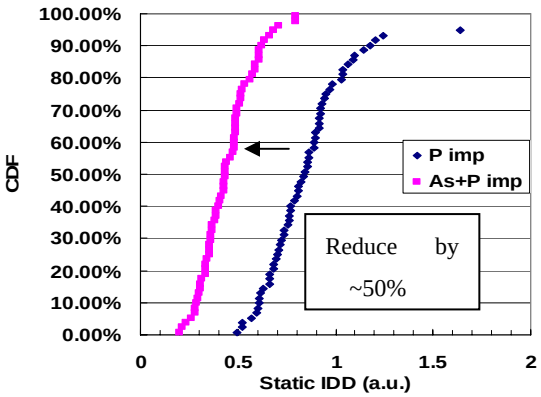


Fig.8 Chip static Idd cumulative plot

Fig. 7 shows the SVT PMOS transistor Id ~ Vg curves with a dimension of width/Lengh=10/0.06um. The dual halo implantation data is from high arsenic dosage split. For Vd=-0.1V, dual halo implantation approach has similar curve with phosphorus only approach. But for Vd=-1.32V, the I-V curve for dual halo implantation approach has a shift to the right hand side as compared to phosphorus only approach. However, the sub-threshold slope is similar for the two approaches, which indicates that dual halo implantation approach has better controllability of drain induced barrier lowering (DIBL). The phenomenon also explains the fact that SVT transistor has lower IOFF for dual halo implantation approach than for phosphorus only as shown in Fig. 6. The narrower the transistor, the more Ioff difference of these two approaches would show, due to that phosphorus out-diffusion ratio increases inversely proportion to the transistor width for phosphorus only approach.

Further more, dual halo implantation approach can improve Gm*Rout by 20% as compared to the phosphorus only approach. The Gm*Rout gain is mainly due to the Rout increase as a result of dual halo implantation has better DIBL controllability. This Gm*Rout gain would particularly benefit analog circuits.

Dual halo implantation approach can improve INWE and reduce SVT transistor off

current, so it reduces digital product chip standby current, as shown in Fig. 8. Dual halo implantation approach reduces our digital product chip static IDD by 50% compared with phosphorus only approach. Further more, dual halo implantation approach gives tighter cumulative distribution of static IDD.

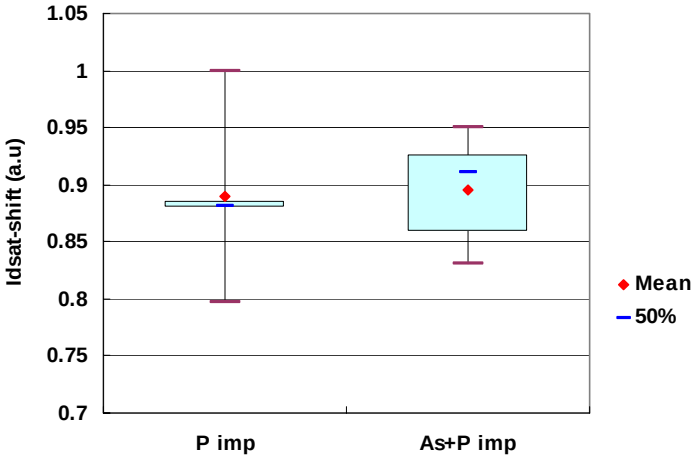


Fig. 9 PMOS Idsat shift under 4000s NBTI stress

For reliability performance, dual halo implantation approach has comparable NBTI performance with phosphorus only approach. Fig. 9 illustrates that these two approaches have comparable Idsat shift under 125C, 4000s, and same Vg stress.

Conclusion

Device with shallow trench isolation (STI) always has inverse narrow width effect (INWE). The conventional methods of improve INWE is focused on STI module optimization, such as active area corner rounding, step height and divot control, etc. In this paper, dual halo implantation approach with arsenic implant and phosphorus implant is utilized to improve INWE due to less arsenic out diffusion into STI oxide, and it gains larger Gm* Rout, lower digital product chip stand by current, and no reliability degradation.

Acknowledgements

The useful reliability related discussion with my colleagues, Kai Zheng, Zhenghao Gan, is acknowledged.

References

[1] Kaushik Roy, Saibal Mukhopadhyay, and Hamid Mahmoodi-Meimand, *Proceedings of The IEEE*, Vol. 91, p305 (2003).

[2] S.Matsuda, T. Sato, H. Yoshimura and etc, *IEDM*. p.137(1998)

ECS Transactions, 18 (1) 111-116 (2009)
10.1149/1.3096437 © The Electrochemical Society

A Novel Super-Halo Doping Concept for Nanoscale CMOS Technology

Congwei Liao[a,b], Shengdong Zhang[b]
[a] Institute of Microelectronics, Peking University, Beijing 100871, P. R. China
[b] Shenzhen Graduate School, Peking University, Shenzhen 518055, P. R. China
Email: zhangsd@pku.edu.cn

A novel super-halo doping concept for nano-scale CMOS application is presented, which features a counter ion implantation performed through the gaps formed at the gate edges. The implantation forms two high localized halo doping (LHD) zones below the source and drain regions adjacent to the channel. The LHD zones effectively shield the source and channel regions from the drain field penetration, ensuring a much improved short-channel effect. The excellent performances of the LHD devices are verified by simulation results. The fabrication process for the LHD devices is also suggested and discussed.

1. Introduction

As the gate length of CMOS devices is scaled down to sub-20nm regime, many problems become difficult to solve in the conventional bulk CMOS. Compared with the bulk CMOS, silicon-on-insulator (SOI) CMOS technology promises the better scalability in addition to offering stronger driving ability and better isolation between devices. Two types of SOI devices, namely the partially depleted SOI (PDSOI) and fully depleted SOI (FDSOI) have been extensively investigated so far. The short channel effects (SCEs) can be controlled well in the FDSOI device if the silicon film is ultra thin. However, the threshold voltage of the UTB device is very sensitive to the UTB thickness, resulting in a critical requirement for UTB film thickness uniformity, which is extremely difficult to meet with the current SOI techniques [1]. It is thus unlikely to introduce the UTB SOI devices into ULSI in the near future. On the other hand, the PDSOI device has been considered suffering from the worse floating body effect and short channel effect and thereby being not a good candidate for ULSI application. The floating body effect has not been a major issue any longer as the supply voltage has scaled to around 1 V, however, the SCE with the PDSOI device remains a major concern.

In general, the PDSOI device is very similar to the bulk one in the design for depressing SCEs. The retrograde and halo (pocket) doping, two typical body doping ways of the nano-scale bulk MOSFETs, can also be applied in the PDSOI device [1]. The retrograde doping introduces a vertically non-uniform doping profile with the

Ex. 1016, Page 139

light doping at the surface and the heavy doping in the deeper region [2-9]. The light doping is for attaining the low threshold voltage and high mobility, and the heavy doping for preventing S/D punch through. However, the retrograde doping usually leads to the formation of the p^+/n^+ junction both at the side and bottom of the source/drain region. The p^+/n^+ junction brings a significant parasitic capacitance and leakage current. On the other hand, the halo doping implemented by a tilted implantation is able to eliminate the formation of the p^+/n^+ junction at the bottom, but causes a serious doping compensation in the S/D region, which results in a significantly increased parasitic resistance. And, the halo doping usually needs to be carefully designed to not induce a reverse short channel effect. For the extremely scaled devices, the super-halo doping profile with the dopant concentration non-uniform both vertically and laterally is required [9]. However, the super-halo doping profile is too hard to implement in present techniques.

In this paper, a novel super-halo doping approach called high localized halo doping (LHD) is proposed. The PDSOI devices with the proposed super-halo doping are studied with device and process simulation tools. Results show that the PDSOI device with the LHD doping could be scaled down well to 20 nm and below. And a feasible fabrication process compatible with the current CMOS technology for implementation of the LHD doping is designed.

2. Scalability of PDSOI Devices

In order to investigate the scalability of PDSOI devices, the short channel effects of the devices with uniformly doped (UD) channel are firstly studied with a device simulator. In the simulation, the extension junction depth is 10 nm, the gate oxide

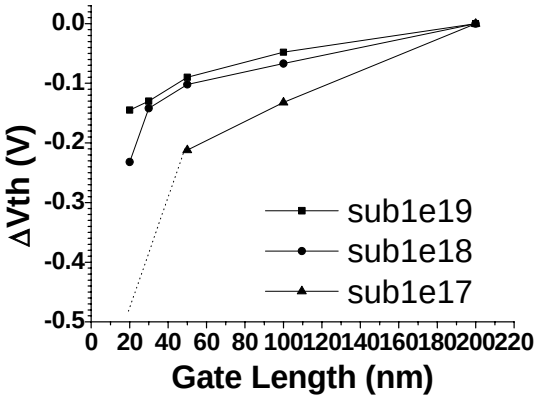


Figure 1. The threshold voltage roll-off (ΔVth) versus the gate length of the UD PDSOI device with different channel doping concentration

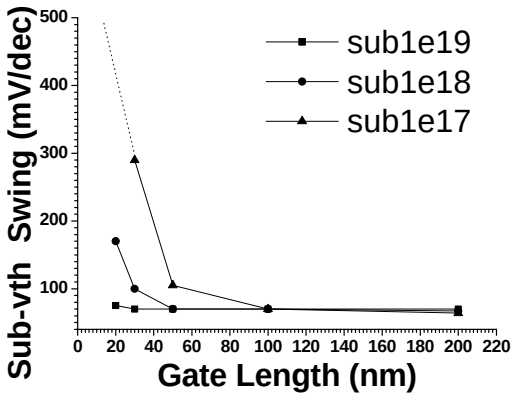


Figure 2. The sub-threshold swing (SS) versus the gate length of the UD PDSOI device with different channel doping concentration

thickness is 1 nm, and the thickness of SOI and BOX are 40 nm and 40 nm, respectively. Figure 1 shows the threshold voltage roll-off of the UDPDSOI device. It

Ex. 1016, Page 140

is observed that the threshold voltage roll-off can be alleviated well if the channel doping concentration is up to more than 10^{19} cm^{-3}, where the acceptable Vth variation of smaller than 150 mV can be achieved in the sub-20 nm devices. That the heavy channel doping helps to reduce Vth roll-off is attributed to the less charge sharing between source/drain and channel. Figure 2 shows the dependence of the sub-threshold swing (SS) on the dopant concentration in the channel. It is seen that the sub-20 nm PDSOI devices with the acceptable SS of smaller than 90 mV/dec can be implemented if the channel doping concentration is up to be around $10^{19}/cm^3$.

Although the heavy channel doping can suppress the SCEs, it results in the deteriorated carrier mobility, large parasitic resistance, large parasitic capacitance and significant tunneling current. Figure 3 shows the dependences of the effective carrier mobility on the doping concentration in the channel, indicating noticeable mobility deterioration when the dopant concentration is high enough.

In the bulk CMOS technology, in order to effectively alleviate the undesired effects caused by the heavy doping, the super halo doping (SHD) is usually used to implement the sub-25nm bulk CMOS device [9]. The investigation demonstrates that SHD provide the sub-20 nm device with both much suppressed SCEs and high

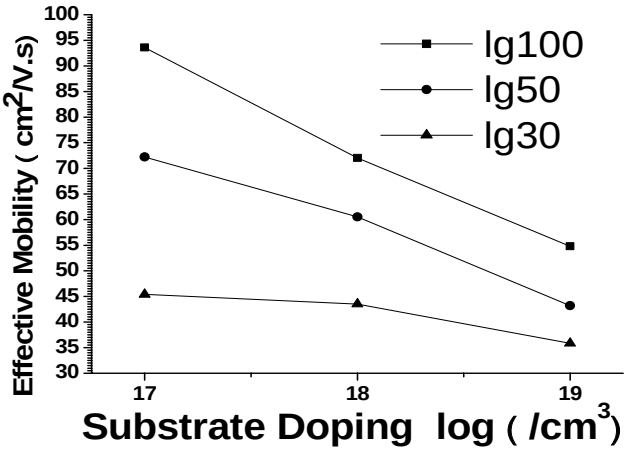


Figure 3. Effective mobility versus doping concentration of UD PDSOI at Vgs-Vth=0.35V and Vds=0.3V

driving ability. As the PDSOI device is much similar to its bulk counterpart in the device structure, the SHD method is believed to be also applicable to the sub-20nm PDSOI device. However the SHD is too hard to implement by the conventional technology. Therefore, more feasible approaches to SHD are highly desired.

3. Concept and Implantation of LHD

In this work, we propose a novel way to implement the SHD, which is called Local Halo Doping (LHD). Different from the conventional halo doping which is

implemented through tilt implantation after the gate patterning, LHD is implanted through gaps formed at the gate edges.

Figure 4 shows a CMOS compatible fabrication process for implementing the proposed LHD concept. The process starts with the conventional CMOS technology to form the gate dielectric, gate electrode and gate oxide cap as shown in Figure 4(A). Then a sacrificed nitride and sacrificed oxide (II) layers are deposited successively and planarized by chemical-mechanical polish (CMP) to expose the nitride as shown in Figure. 4(B). After the nitride etching in hot H_3PO_4, two gaps at the gate edges are formed as shown in Figure. 4(C). Through the gaps, the ion implantation is performed,

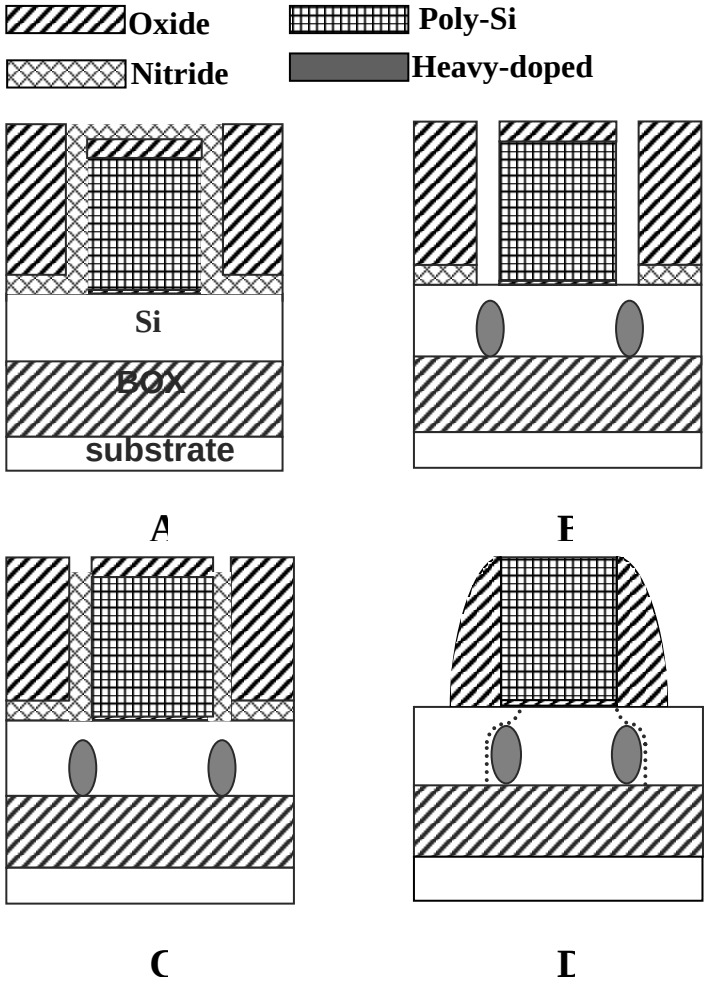


Figure 4. Major steps of the proposed fabrication process for the LHD PDSOI transistor

forming a high localized counter-doping zone in the source/drain region adjacent to the channel region as shown in Figure. 4(D). After removing the sacrificed oxide and nitride, the process turns to the conventional back-end steps, including ion implantation for source/drain doping, rapid thermal annealing (RTA), silicidation and so on.

In the proposed LHD method, the serious source/drain doping compensation is

Ex. 1016, Page 142

avoided, and the highly counter doping region underneath the source and drain extension region is favorable for suppressing SCEs.

4. Device Performances of PDSOI Device with LHD

The performances of the sub-20nm PDSOI devices with the novel proposed LHD are studied using the device and process simulators. In this simulation, the gate thickness is 1nm and the thickness of SOI and BOX are 40nm and 40nm respectively. For comparison purpose, the conventional retrograde doping (RD) PDSOI device is also simulated.

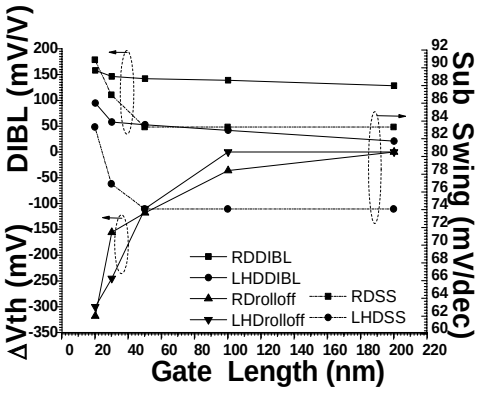


Figure 5. Drain induced barrier lowering (DIBL) effect, threshold voltage roll-off (ΔVth) and Sub-threshold swing (SS) versus the gate length with the RD PDSOI and LHD PDSOI transistor.

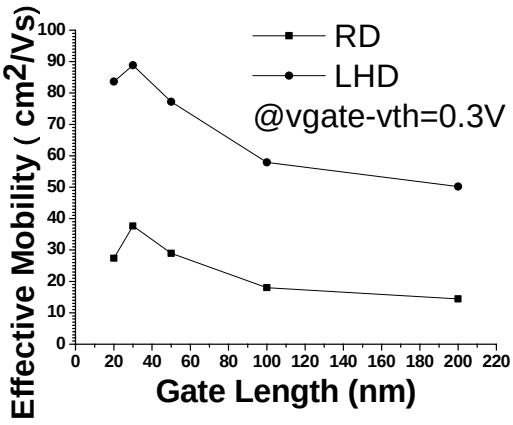


Figure 6. Effective mobility versus Lg with RDPDSOI and LHDPDSOI devices at Vgs - Vth = 0.3 V and Vds = 1.0 V

Figure 5 shows the threshold voltage roll-off (ΔVth), the DIBL effect and the SS versus Lg with the LHD PDSOI and RD PDSOI devices. It is observed that for suppressing of threshold voltage roll-off (ΔVth), the LHD device is the same as the RD device. This is attributed to the same HDD junction depth and the same surface doping concentration. However, in the suppression of DIBL, the LHD device is stronger than the RD device. The better DIBL with the LHD PDSOI device is due to the lateral non-uniform channel doping. Furthermore, it is observed that in the Sub-threshold Swing (SS), the LHD device is better than the RD device. The much improved SS of LHD PDSOI device owes to the less channel doping concentration. Therefore, the LHD PDSOI device possesses the stronger short-channel effect immunity than its RD PDSOI counterpart.

Figure 6 shows the effective mobility versus Lg with RD PDSOI and LHD PDSOI devices. The effective mobility is extracted at the bias condition of Vgs-Vth=0.3V and Vds=1.0V. It is observed that LHD PDSOI provides 2 times the effective mobility of the RD device. In other words, the LHD PDSOI device has higher driving ability than the RD PDSOI device. The much improved effective mobility of the LHD PDSOI device is attributed to the less carrier scattering because

Ex. 1016, Page 143

of the reduced channel doping concentration. The reduced channel doping is believed to reduce the vertical electrical field, therefore the carrier scattering is reduced in the LHD PDSOI device. It is also observed that with the gate length scaling down, the effective carrier mobility is increasing first and then decreasing at the 20 nm node. The increased effective mobility with the gate length scaling down is attributed to the reduction of the effective vertical electric field. However, the carrier velocity is limited by the saturation effect for a nanometer channel length which accounts for the effective mobility reduction at the 20 nm node.

6. Conclusions

A new doping method called “localized halo doping, LHD” has been investigated for the first time. In comparison with the traditional halo and retrograde doping methods, the LHD has several attractive merits, such as much suppressed short channel effects and improved effective mobility. At the same time, a CMOS compatible fabrication process for implementing the proposed LHD profile has also been designed. The results show that the LHD approach is a promising doping method applicable to sub-20nm CMOS technology.

Acknowledgments

This work is supported by NSFC of China under Project 60676023 and 60736030, and a foundation for the author of national excellent doctoral dissertation (FANEDD) of China under Project 200440.

References

1. G. G. Shahidi, *IBM journal of research and development*, 46, p.121-123 (2002).
2. D. J. Frank and Y. Taur, *Solid-State Electronics*, p.315-320 (2002).
3. N. E. B. Cowern, A. J. Smith and B. Colombeau, *IEDM*, p.4 (2005).
4. H. Wakabayashi, M.Ueki and M. Narihiro, et al, *IEDM*, p. 49-52 (2000).
5. Bin Yu, Wann, et al, *IEEE Transactions on Electron Devices*, 44, p. 627-634 (1997).
6. Y. Toyoshima, K. Miyashita, and E. Morifuji, *Junction Technology*, p.41-44 (2000).
7. Chun-Hsing Shih ,Yi-Min Chen and Chenhsin Lien, *IEEE Transactions on Electron Devices*, 12, p.39- 42 (2003).
8. Chun-Hsing Shih ,Yi-Min Chen and Chenhsin Lien, *IEEE Transactions on Electron Devices*, 11, p.2294- 2297 (2003).
9. Y. Taur, C. H. Wann, and D. J. Frank, *IEDM*, p.789-792 (1998).

ECS Transactions, 18 (1) 117-122 (2009)
10.1149/1.3096438 © The Electrochemical Society

Shallow Trench Isolation Stress Effect on NMOS Transistor Leakage, SRAM Standby Current and V_{CCMIN}

Jianhua Ju*, Eric Liu, ZhaoXu Shen, Allan Zhou, Jinhua Liu, Daniel Deng, Cathy Ren, Susu Wei, Hokmin Ho and Jay Ning

Logic Technology Development Center, SMIC (BJ) Beijing 100176, P. R. of China
*E-mail: Jianhua_Ju@smics.com

Three key process steps in shallow trench isolation (STI) technology: STI liner oxide, STI to AA step height, and annealing temperature. Their impacts on NMOS devices leakage, SRAM standby current and Vccmin have been studied. MOSFET transistors off state current (Ioff) and body current (IB) were measured at varied length of diffusion (LOD) for STI stress effect analysis. For low power NMOSFET transistors, IB is seen to be the main part of Ioff and has a strong correlation to the processes that modulates STI stress. The experiment results show that the STI stress level can be much reduced by decreasing STI oxide RTA temperature and liner ox thickness, hence leading to lower IB and SRAM standby current. It is also demonstrated that SRAM Vccmin can be improved ~10% by STI stress reduction.

Introduction

Shallow trench isolation (STI) induced elastic stress effect on MOSFET transistor becomes ever more significant as the transistor dimension scales down and has been widely studied (1-3). NMOS Idsat/off performance will be much degraded by STI stress as the STI stress may enhance boron diffusion in the channel region. This ehanced boron diffusion leads to a thinner depletion region between source-drain extension and channel and increases the band to band tunneling (BTBT) junction leakage (4,5). The STI stress induced boron diffusion is more serious for small dimension devices, and may cause severe threshold voltage Vt mismatch and worsen SRAM Vccmin (5).

In this work, STI liner ox thickness, STI ox anneal temperature and step height effects on MOSFET transistor's junction leakage have been investigated by measuring IB with different LOD area. The small devices junction leakage and SRAM standby current and Vccmin cumulative distribution have also been studied. This study will be much help for devices leakage reduction and SRAM Vccmin improvement.

Ex. 1016, Page 145

Experimental Detail

The wafers are prepared with 65nm low power logic technology process flow. The samples are fabricated with 300mm P type (100) epi-silicon wafers with resistance between 8 to 12 ohm/cm. Different STI liner oxide thickness, STI step height and STI oxide anneal temperature as shown in table 1 have been designed to study STI process effects on electric parameters of small device and SRAM cell.

Table 1 STI Experimental Condition

Sample ID	Experimental Condition
SH RTA	STI HDP high temperature annealing
LT RTA	STI HDP low temperature annealing
SH BL+50 A	STI ox strip time BL – 60 s
SH BL-50 A	STI ox strip time BL + 60 s
Thin Liner	STI thinner liner ox
Thick Liner	STI thicker liner ox

In the experiments, STI stress impacts on devices body current IB are monitored through a set of test structures with different diffusion (LOD) length as shown in Figure 1. For low power MOSFET transistors, the BTBT junction leakage from source/drain extension to channel is the main part of Ioff. According to (1), IB behaviour can be used to study the junction leakage. Figure 2 shows IB *vs*. LOD and width. It is seen that the IB scales down with LOD and W, indicating STI stress enhances boron diffusion hence leads to device junction leakage increasing for smaller dimension LOD and W.

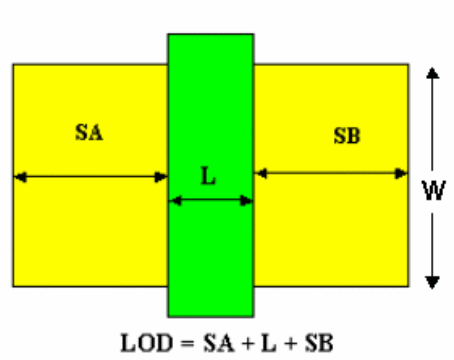
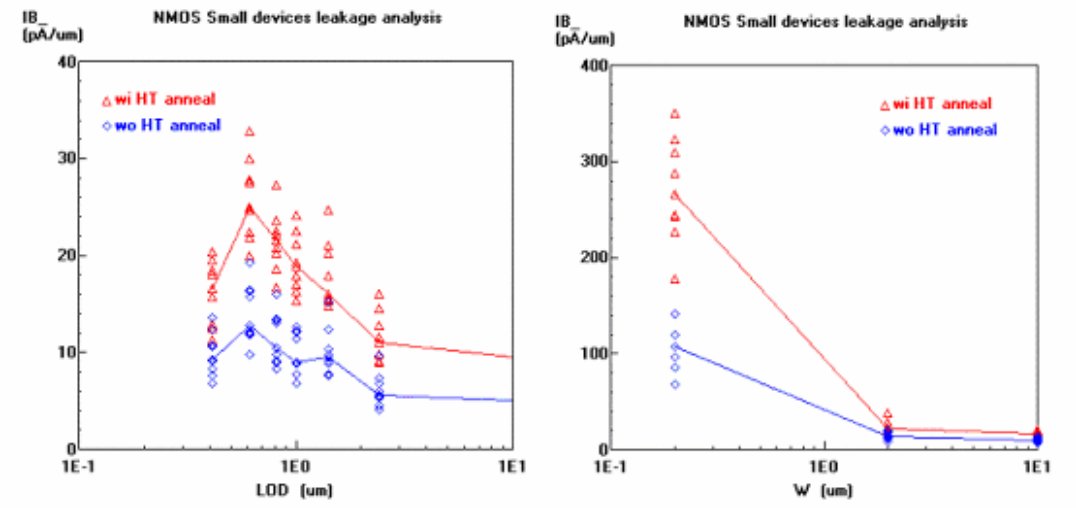


Figure 1 Schematic MOSFET transistor showing the definition of LOD

Ex. 1016, Page 146



(a) (b)

Figure 2 NMOS body current IB variations with LOD (a) and channel width (b)

Results and Discussion

STI HDP oxide refilling post high temperature annealing will cause compressive stress, which enhances boron diffusion and leads to higher IB as shown in figure 2 (a). Moreover, it is confirmed that the change in IB is more sensitive to stress variations when LOD and/or the device width are decreasing. The result indicates that small devices IB can be used to monitor the stress variation during the STI process and the STI stress impact on SRAM standby current and Vccmin.

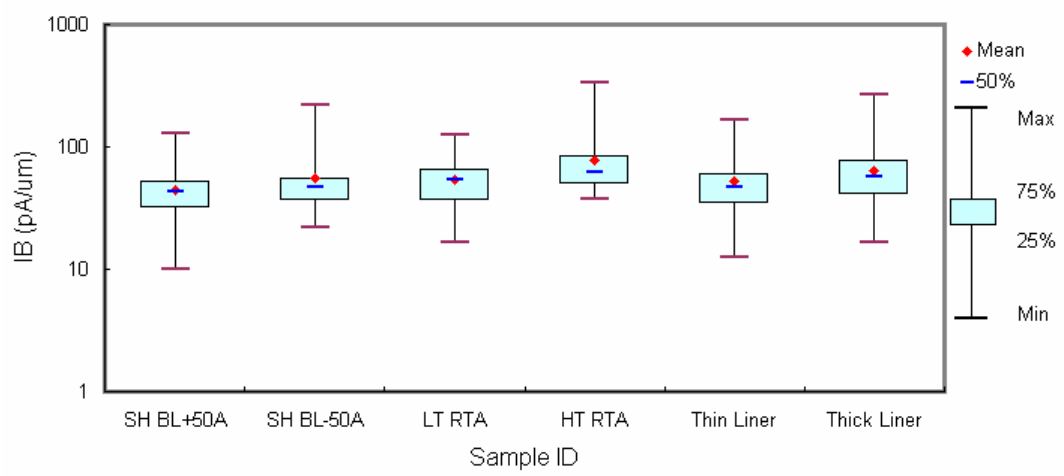


Figure 3 Small devices (W/L=0.12um/0.06um) IB for different STI experiment

As shown in Figure 3 the small NMOSFET transistor (W/L=0.12um/0.06um) IB will be increased with STI step height reduction. IB is also shown to be increased by enhancing STI oxide RTA temperature or increasing STI liner oxide thickness. Obviously, all these three STI processes will lead to STI compressive stress

Ex. 1016, Page 147

increasing and this compressive stress enhances NMOSFET transistor source-drain extension (SDE) to channel BTBT junction current (3).

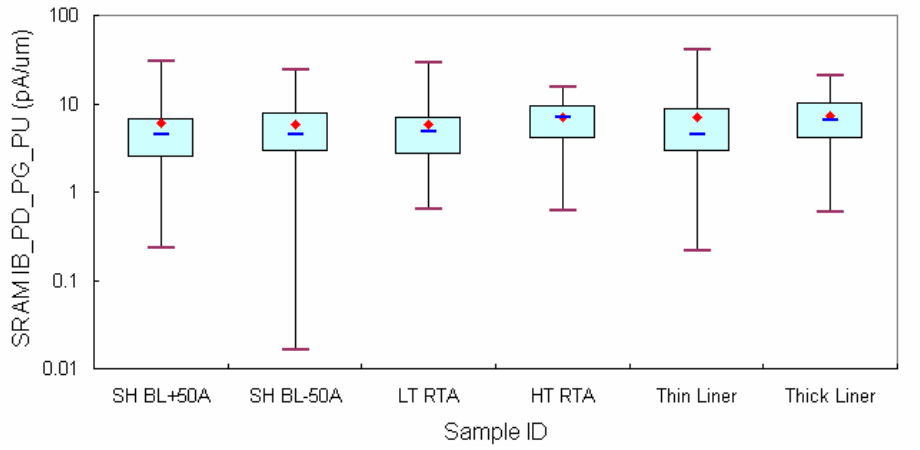


Figure 4 SRAM devices IB for different STI process experiment

The same results can be obtained from total IB of the SRAM cell as shown in Figure 4. The higher the compressive stress of the STI process, the higher the observed junction leakage of the SRAM devices. Moreover, SRAM standby current correlation with STI step height and liner oxide thickness, as shown in Figure 5, has the same trend with that of IB.

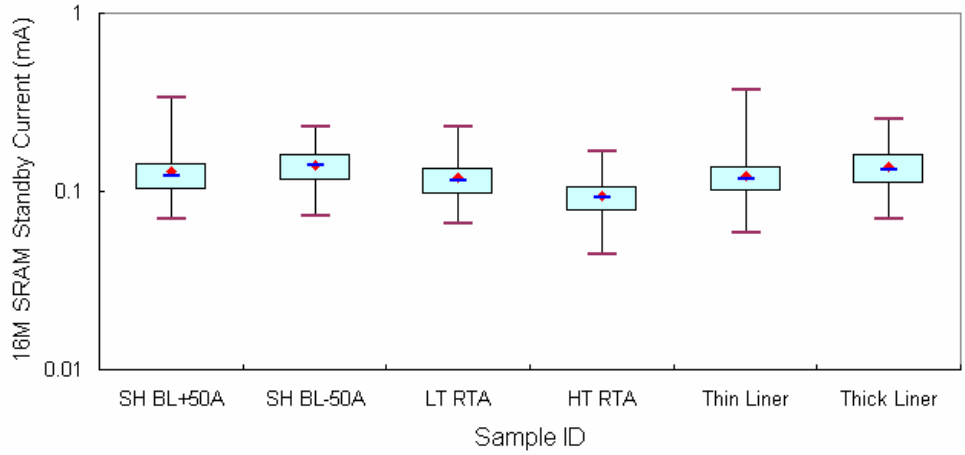
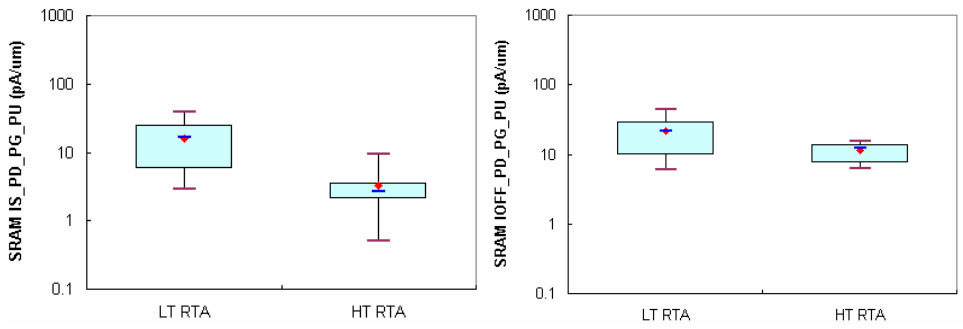


Figure 5 16M SRAM standby current for different STI process experiment

However, though STI stress is supposed to be smaller for low STI filling post anneal temperature than that of high temperature, SRAM standby current of low STI filling post aneal temperature split is shown to be higher than that of high temperature This is because SRAM standby current depends on the Ioff of its cell devices (pull up, pull down and pass gate), IB is just one part of the Ioff. As shown in Figure 6, low anneal temperature has lower IB but has higher IS and Ioff than high temperature anneal device.



(a) (b)

Figure 6 (a) SRAM cell subthreshold leakage and (b) off state current

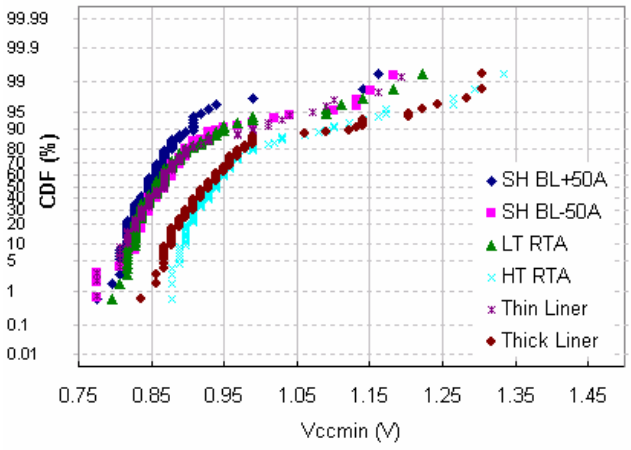


Figure 7 the cumulative distribution of 16M SRAM Vccmin for different STI process experiment

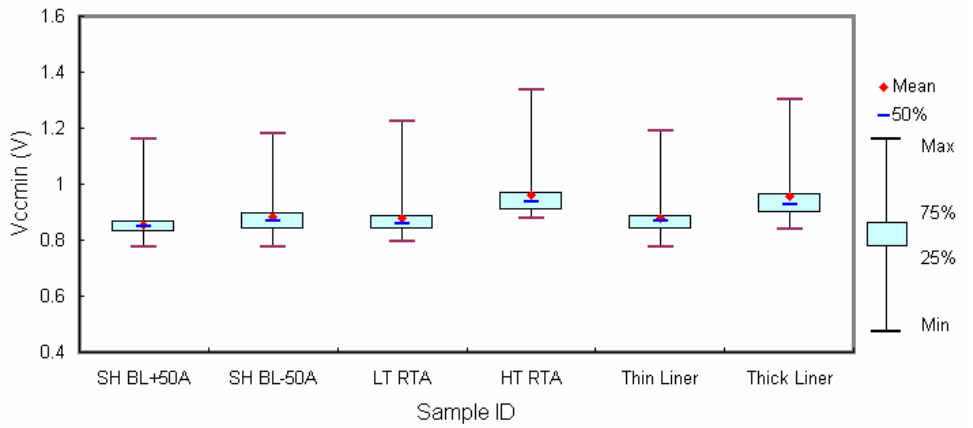


Figure 8 The Vccmin of 16M SRAM array for different STI process experiment

Vccmin cumulative distribution of SRAM has been studied as shown in Figure 7 and Figure 8. It is seen that Vccmin has strong correlation with STI stress. The higher the STI stress is, the higher SRAM Vccmin value will be. The result indicates that higher STI stress will induce higher SRAM Vt mismatch. It is also shown in Figure 8 that, Vccmin of LT RTA and Thin Liner processes are about 90mV lower than

Ex. 1016, Page 149

that of HT RTA and Thick Liner processes, which means that both low temperature anneal and thin liner processes can improve Vccmin as much as 10%.

Conclusion

It is shown in the paper that by decreasing STI oxide RTA temperature, reducing liner ox thickness or increasing STI step height, the STI compressive stress level is much reduced and device IB and SRAM standby current can be much reduced also. Besides this, SRAM Vccmin is shown to be improved ~10% by reducing STI stress.

Reference

1. Hsing-jen Wann et al., *IEDM Tech. Dig.*, P147-150, 1992
2. Venkatesh Gopinath et al., *ESSDERC* P631-633, 2002
3. Yi-Ming Sheu, Sheng-Jier Yang et al., *IEEE Transactions on Electron Devices*, Vol. 52 No.1 2005 PP.30-38.
4. Masafumi Miyamoto *IEEE Transactions on Electron Devices*, Vol. 51 No.3 2004, pp 440-443
5. Tung Hsing Lee et al, *J. Phys. D: Appl. Phys*. 41 195101 (2008)

ECS Transactions, 18 (1) 123-128 (2009)
10.1149/1.3096439 © The Electrochemical Society

A High Performance 80V Smart LDMOS Power Device Based on Thin SOI Technology

Gene Sheu [a], Shao-Ming Yang [a], and Yu-Shan Hsu [a]

[a] Computer Science and Information Technology (CSIE) Department of Asia University, Taichung County, 41354,Taiwan

The high junction leakages, circuit latched issues, and high parasite capacitances preset in junction isolation technology can be improved by the thin SOI (Silicon-On-Insulator) technology. A CMOS compatible SOI technology will be in of the technologies used in the future roadmap of LDMOS devices. A CMOS compatible thin SOI LDMOS (Lateral Double-diffused MOSFET) device, with 0.18 micron gate length, 0.02 micron gate oxide and 3 micron N-drift region, is proposed to achieve the optimal (BVoff) off-state breakdown voltage and on-state resistance (Ron) values. The characteristics of the proposed LDMOS device are verified by the two-dimensional process simulator TSuprem-IV and the device simulator Medici. The simulated results have shown that a device performance at the range of BVoff, 80V, and Ron, 190 mohm-mm2, is achieved. The on-state breakdown voltage is measured at 70V with an excellent safe operating area (SOA) performance for the drain source current versus on-state breakdown voltage.

INTRODUCTION

Due to increasing system demands, the need to integrate multiple functions (e.g. digital, analog, high voltage, etc.) on a single chip has become a trend. In the field of the "Smart-Power" integrated circuits (ICs), traditional integration approaches have focused on the optimization of high-voltage transistors to reduce device size. By using the SOI technology, it will improve the problems such as device latch up, leakage current, parasitic capacitance, and oxide isolation which reduces the threshold voltage shift caused by radiations. Since a LDMOS device is designed with a drift region between the gate and the drain, it will stand a high voltage operation compared with CMOS. Therefore, BCD devices will implant the LDMOS structure for high voltage operations.
A SOI of 3um epi is used due to low cost to make a 80V LDMOS by the BCD compatible processes proposed in this paper. Therefore, CMOS for lower power logics, Bipolar for high current amplifications and LDMOS for the high voltage applications can all be integrated in the proposed processes. BCD devices are widely used for motor, automotive, display and digital audio applications.

Ex. 1016, Page 151

DEVICE STRUCTURE AND PROCESS

LDMOS Device Structure and Process

The SOI LDMOS device structure is shown in Figure 1: 2 um buried oxide (BOX), 3 um Epi layer, and 200 Å gate oxide with Vt = 1.8V. This device will achieve 80v breakdown voltage, and 1.952 $m\Omega.cm^2$ on-state resistance.

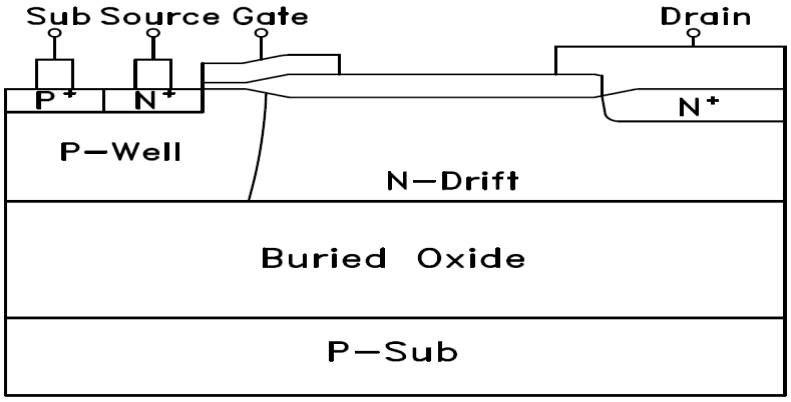


Figure 1 LDMOS Structure

The length of N-drift region can impact on the device breakdown voltage so that the N-drift region of 3 um is adapted to achieve the requirement of more than 80V breakdown voltage in this study. The electrical field distributions near the gate and the drain will depend on the dose densities of P-well and N-drift region so that the optimal breakdown voltage can be achieved only by intensively multiple simulations of device parameters.

CMOS Device Structure and Process

The CMOS device structure is shown in Figure 2: 2 um buried oxide (BOX), 3 um Epi layer, and 200 Å gate oxide with Vt = 1.8V. This device will achieve 5v breakdown voltage.

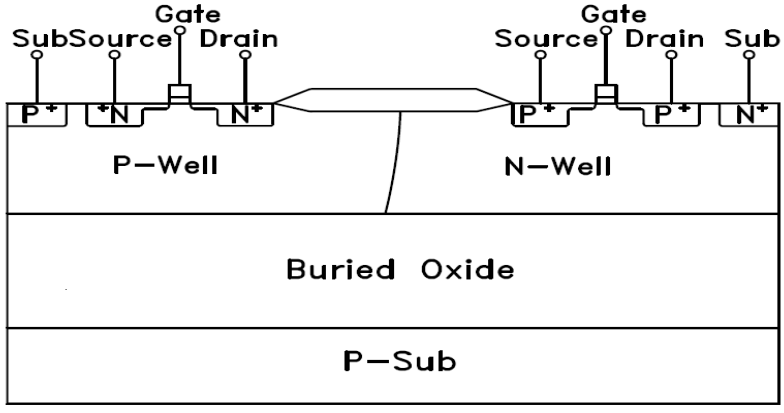


Figure 2 CMOS Structure

Based on the minimum changes in the traditional CMOS process to achieve the compatible process for LDMOS devices, the source and drain of 0.18 um channel CMOS must be fabricated and accomplished without punch-through after LDMOS completes the drain and source, and the annealing of P-well and N-drift process.

RESULTS AND DISCUSSION

The simulation results of LDMOS

The electrical potential and electrical field distributions of LDMOS devices are simulated by MEDICI. In Figure 3 the electrical potential distribution along the N-drift region is very linear so that the electrical field distribution along the N-drift is very uniform and flat as shown in Figure 4, and is good for breakdown voltage; the electrical field distributions in N-drift region usually peak near the drain side or source side or both sides if device parameters are not optimized.

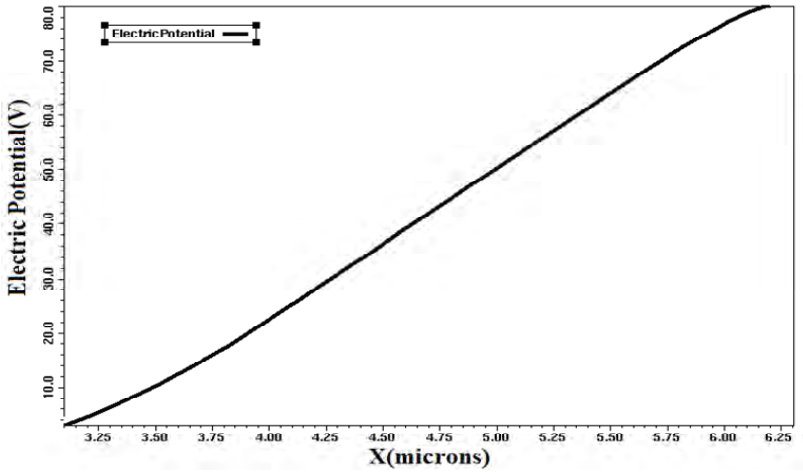


Figure 3 LDMOS surface electrical potential of LDMOS along N-drift region

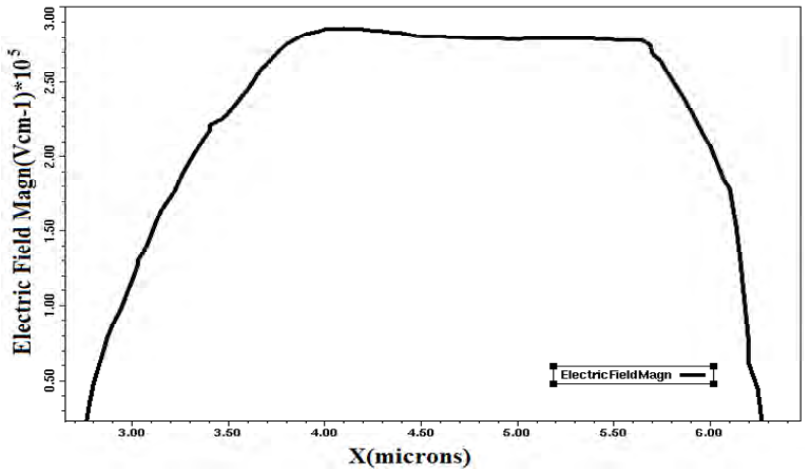


Figure 4 LDMOS surface electrical field of LDMOS along N-drift region

The I-V curves of LDMOS with gate voltages of 2,4,6,8, and 10V are shown in Figure 5, and the safe operating area (SOA) of breakdown voltages in the Figure 5 is shown in Figure 6.
The on-state breakdown voltages of LDMOS with gate voltages of 2,4,6,8, and 10V are 92, 75, 64, 58, and 54V respectively. There are no kink effects shown in Figure 5.

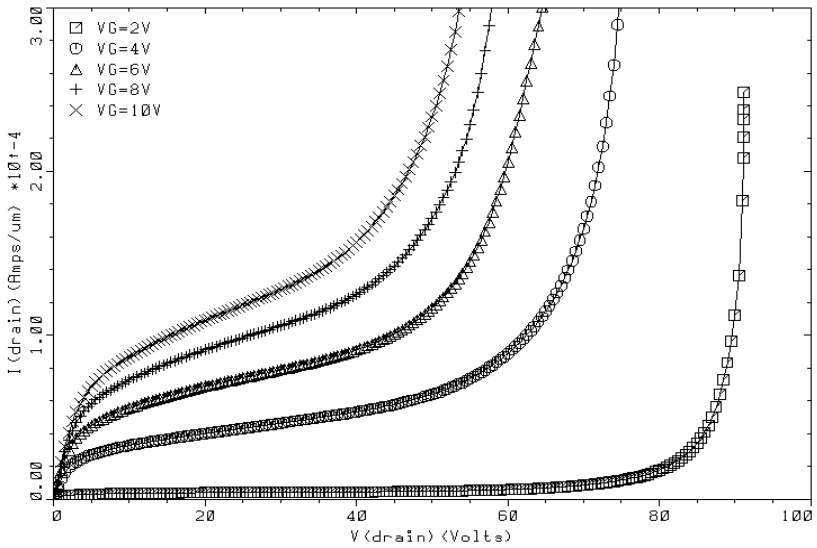


Figure 5 I-V curves of LDMOS with different gate voltages.

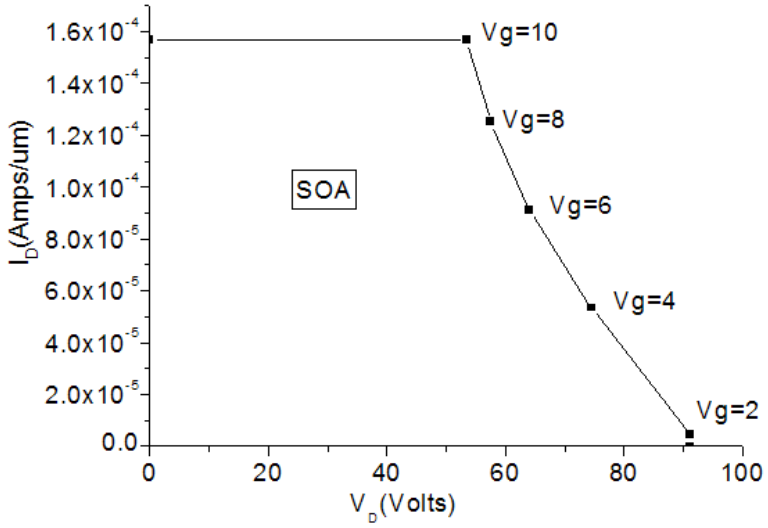


Figure 6 Safe operating area of LDMOS breakdown voltages

The simulation results of CMOS

The CMOS device made by the proposed processes has shown good I-V characteristics of NMOS as shown in Figure 7. However, the I-V characteristics of PMOS shown in Figure 8 are not as good as that of NMOS, and behave like a resistor.

Ex. 1016, Page 154

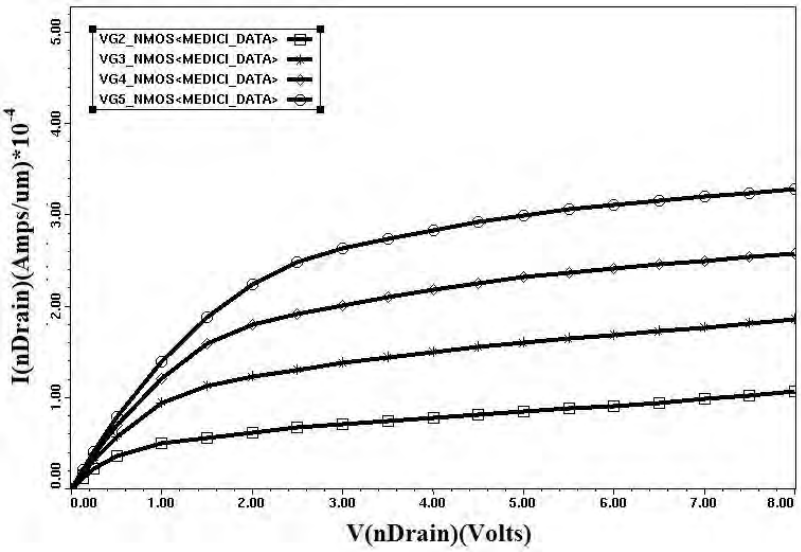


Figure 7 I-V characteristics of NMOS with different gate voltages.

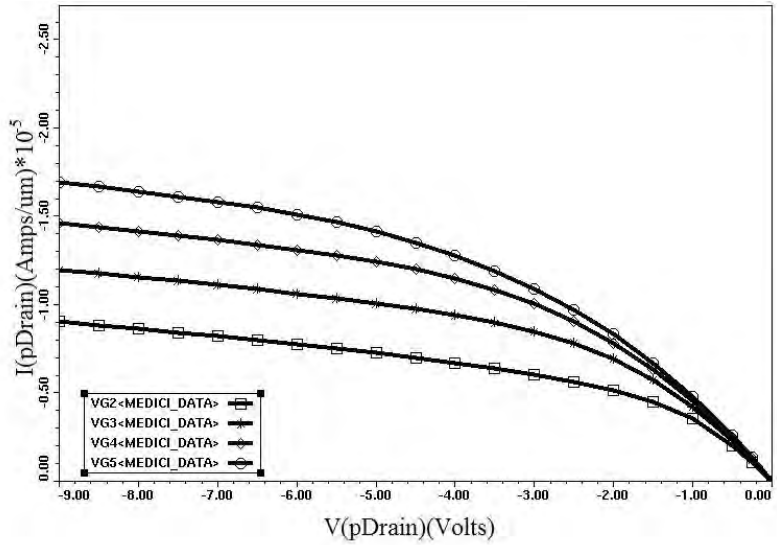


Figure 8 I-V characteristics of PMOS with different gate voltages.

CONCLUSION

The SOI LDMOS devices made by the proposed processes have achieved good off and on state breakdown voltages. The SOA of breakdown voltages is also obtained. The simulated results of CMOS devices processed along with LDMOS devices have shown good I-V characteristics of NMOS. Although I-V characteristics of PMOS behave like a resistor, we believe the proposed processes for BCD devices are feasible if the drain and source of CMOS devices are fabricated after LDMOS completing the drain and source, and the annealing of P-well and N-drift process.

Ex. 1016, Page 155

References

1. Choul-Joo Ko, Sang-Yong Lee, Il-Yong Park, Cho-Eung Park, Bon-Keun Jun, Yong-Jun Lee,Chan-Hee Kang, Jae-O Lee, Nam-Joo Kim, and Kwang-Dong Yoo, Implementation of 85V High Side LDMOS with n-layer in a 0.35um BCD Process,2008 IEEE,p.103-p.106,(2008).
2. T. Nitta, S. Yanagi, T. Miyajima, K. Furuya, Y. Otsu, H. Onoda, and K. Hatasako, Wide voltage power device implementation in 0.25um SOI BiC-DMOS, IEEE,p4-8(2006).
3. Benno H. Krabbenborg & Jacob A. van der Pol, Robustness of LDMOS power transistors in SOI-BCD processes and derivation of design rules using thermal simulation, International Symposium on Power Semiconductor Devices, IEEE,p.157-160 (2001).
4. Paul G.Y. Tsui, Percy V. Gilbert, and Shih Wei Sun, Integration of Power LDMOS into a Low-Voltage 0.5 um BiCMOS Technology,1992 IEEE, p. 2.3.1-p. 2.3.4,(1991).
5. S.M.Sze, Physics of semiconductor devices 2nd edition, p. 469,Taiwan(1981).
6. Qiao Ming,, Zhou Xianda, Duan Mingwei, Fang Jian, Zhang Bo, and Li Zhaoji, Breakdown Characteristic of Multiregion Double RESURF LDMOS with High Voltage Interconnection ,CHINESE JOURNAL OF SEMICONDUCTORS, Vol. 26 No. 10 , p.1428-p1432 (2005).
7. Seung-Yaup Han,long-Min Na,Yearn-lk Choi,lin-Cheol Shin,Sang-Koo Chung, AN ANALYTICAL MODEL OF THE BREAKDOWN VOLTAGE AND MINIMUM EPI LAYER LENGTH FOR RESURF pn DIODES, Solid-Srare Electronics Vol. 39, No.8, pp. 1247-1248, (1996).

ECS Transactions, 18 (1) 129-134 (2009)
10.1149/1.3096440 © The Electrochemical Society

Simulation Details for the Electrical Field Distribution and Breakdown Voltage of 0.15μm Thin Film SOI Power Device

Hsin-Chiang You[a], Yen-Ling Liu[a] , Shyh-chang Tsaur[a] , Gene Sheu[a]
[a] Department of Computer Science and Information Engineering, Asia University,
No. 500, Lioufeng Rd., Wufeng ,Taichung ,Taiwan , R. O. C

The ultra-high breakdown voltage behavior of a thin-film SOI LDMOS device with a variation in the lateral doping (VLD) of drift region is analyzed extensively by TSuprem4 and Medici. The on/off state resistance and breakdown voltage depending on device parameters including drift region length, doping density gradient, and gate field plate length are examined thoroughly. The process figure of merit (FOM) has shown excellent result for devices with the optimum linearly doped drift region ranging from 30μm to 60μm. The safe operating area (SOA) for drain source current versus on-state breakdown voltage for a 60μm drift region is also presented.

Introduction

The high lateral electric field caused by the crowded equipotential voltage lines near the ends of the drift region in the conventional single RESURF LDMOS can be improved by an optimal lateral linear doping drift region reported in the previous papers [1,2,3] . A novel LDMOS device with a variation in the lateral doping (VLD) of drift regions is demonstrated in a thin film silicon-on-oxide (SOI) technology [4, 5]. An optimum doping gradient in the drift region generating a uniform electric field in the drift region can achieve the best breakdown voltage as demonstrated in previous papers [6, 7] for the junction isolated LDMOS device, and its formula to calculate the gradient of a doping profile is proved to be valid by Medici simulations for the thin film (0.15μm) SOI LDMOS with a VLD drift region in this report. The slit width and slit spacing of a drift region mask proposed in the previous papers [8, 9] are used to generate the optimal linearly doped drift region by optimizing the process parameters including the implant dose, implant energy, annealing temperature, and annealing time.

I. Device structure and linear doping implant mask

The detailed geometry for a thin film SOI device is shown in Fig. 1 The schematic of lateral variation doping mask is shown in Fig. 2: X0 ~ X9 (Slip opening), A0~A8 (Slip spacing). The doping profile along the drift region is shown as below: Qd (Doping density :cm^{-2}) , Tepi (Epi thickness), Ld (Drift region length), G (Gradient of doping concentration, cm^{-4}) = 2*(Qd / Tepi) / Ld

Ex. 1016, Page 157

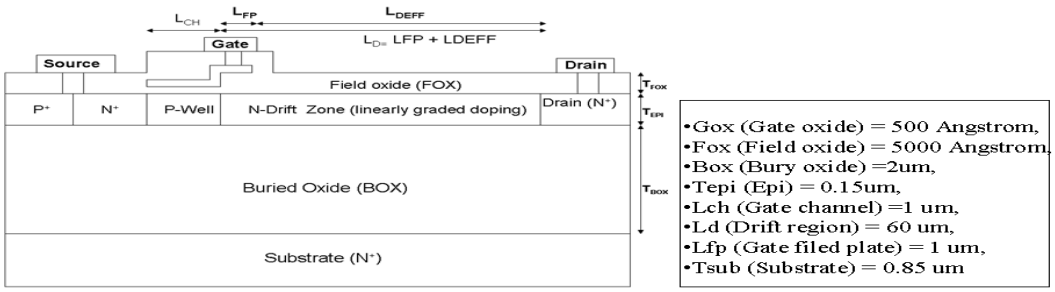


Fig. 1 TFSOI LDMOS schematic cross section

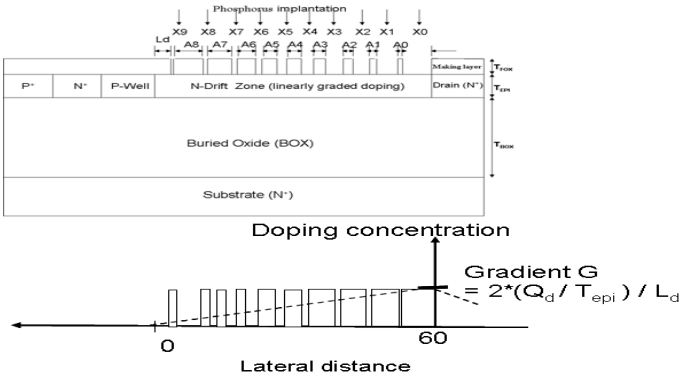


Fig. 2 Cross section of drift region and doping profile along drift region.

The slit opening width and spacing of implant masks for drift region (30um, 40um, 50um, and 60um) are summarized in Table 1:

Design Parameters	Drift Length (30um)	Drift Length (40um)	Drift Length (50um)	Drift Length (60um)	Design Parameters	Drift Length (30um)	Drift Length (40um)	Drift Length (50um)	Drift Length (60um)
X0	4.2	6.6	11.1	13.3	A0	2.5	1.9	1	1.2
X1	2.5	3.7	4.9	5.5	A1	2.5	1.8	1.1	1.3
X2	1.8	2.7	3.5	4.2	A2	2.8	2.2	1.6	1.9
X3	1.4	2.2	3	3.6	A3	3	2.7	2	2.4
X4	0.9	1.7	2.5	3	A4	4.1	2.8	2.4	2.8
X5	0.7	1.2	2	2.4	A5		2.9	2.8	3.4
X6		0.8	1.5	1.8	A6		3.9	2.9	3.5
X7		0.6	1	1.2	A7			2.8	3.4
X8			0.7	0.8	A8			3.9	4.7
X9			0.5	0.6					

Table 1 Lists of the slit opening (x0..x9) and spacing (A0..A8) of implant masks for drift region (30um, 40um, 50um, and 60um) , which were reported in [5].

II. Analytical model.

A schematic cross section of the TFSOI Single RESURF device with a drift region length Ld is shown in Fig. 1, where T_{epi} is the thickness of Si epi layer with dielectric constants ε_{si}, T_{box} is the thickness of buried oxide layer with dielectric constants ε_{ox}, and x, y are the coordinates for measuring the horizontal and vertical position from the gate edge at the interface of the Epi and Field oxide, respectively. The lateral doping profile in a drift region is shown as follows:

$$N_{epi} = G*x + N_0 \qquad [1]$$

Where N_0 is initial epi background concentration and G is the slope of the doping profile. It is assumed that the drift region is completely depleted if the drain, n+n, junction, is applied by a high enough biased positive voltage, and the source, p+n junction, is grounded. The electrical potential φ(x,y) in the Si epi film must satisfy the Poisson equation:

$$\frac{\partial^2 \phi(x,\,y)}{\partial^2 x} + \frac{\partial^2 \phi(x,\,y)}{\partial^2 y} = -\frac{q}{\varepsilon_0 \varepsilon_{si}}(G * x + N_0) \qquad [2]$$

The boundary conditions for the electrical potential are as follows:

$$\phi\,(0,\,y) = 0 \qquad [3\text{-}1]$$

$$\phi\,(\mathrm{Ld},\,y) = \mathrm{Vd} \qquad [3\text{-}2]$$

The continuity of electric displacements are expressed in the following:

$$\left.\frac{\partial \phi(x,\,y)}{\partial y}\right|_{y=0} = -\frac{qN_{fox}}{\varepsilon_0 \varepsilon_{si}} \qquad [3\text{-}3]$$

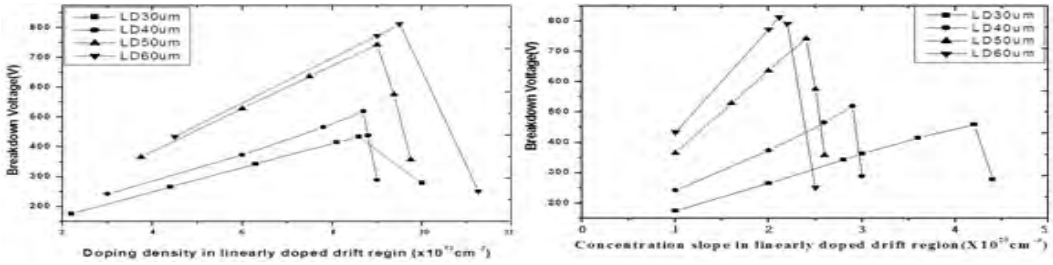
$$\left.\frac{\partial \phi(x,\,y)}{\partial y}\right|_{y=T_{epi}} = -\frac{qN_{box}}{\varepsilon_0 \varepsilon_{si}} - \frac{q\phi(x,T_{epi})}{\varepsilon_0 \varepsilon_{si}} \qquad [3\text{-}4]$$

Where N_{fox} and N_{box} are the interface charge density at the interface of epi-Si/field oxide and epi-Si/buried oxide respectively. The built-in potential and the flat-band voltage are neglected in high voltage devices.

The electric potential φ(x,y), and electric field E(x,y) equations solved based on the above equations are listed in [4].

Results and discussion

Fig. 3(a) shows the dependence of the breakdown voltage on the implant dose density in the drift region. Fig. 3(b) depicts the dependence of the breakdown voltage on the doping profile gradient in the drift region. The optimal breakdown voltage of the devices with drift lengths of 30, 40, 50, and 60um are 459, 519, 743, 810 V, respectively. The simulated results are also consistent with those reported in earlier works [4], [5].



(a) (b)

Fig. 3 The dependence of the breakdown voltage on the implant dose (a) and on the doping gradient slope G (b) for the drift region with lengths of 30, 40, 50, and 60um

Fig. 4 demonstrates the dependence of the on-state resistance on the doping profile gradient in the drift region. The on-state resistance corresponding to the optimal breakdown voltage of the devices with drift lengths of 30, 40, 50, and 60um are 41, 63, 98, 127 (mohm.cm^2), respectively. Fig. 5(a) shows the dependence of the optimal breakdown voltage on the drift region length, and Fig. 5(b) depicts the corresponding doping gradient slope G for the devices with drift lengths of 30, 40, 50, and 60 um.

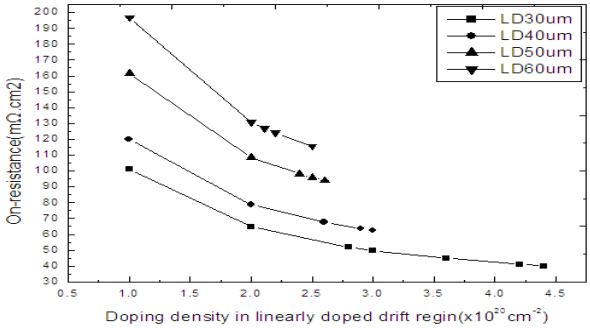
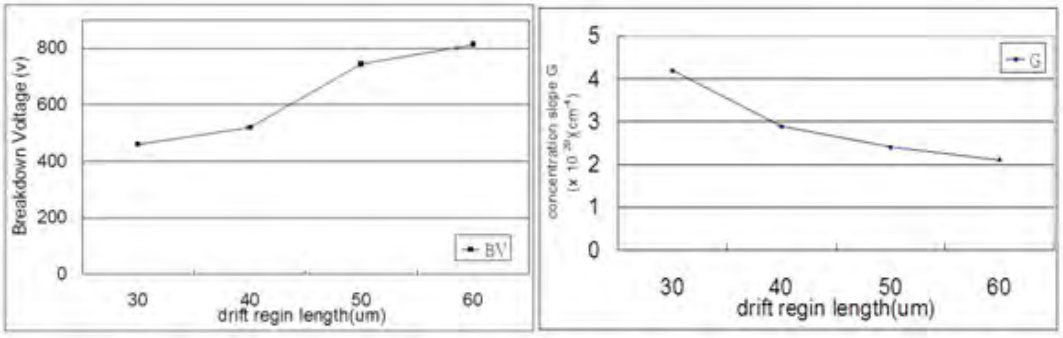


Fig. 4 The dependence of the on state resistance on the doping profile gradient in the drift region.

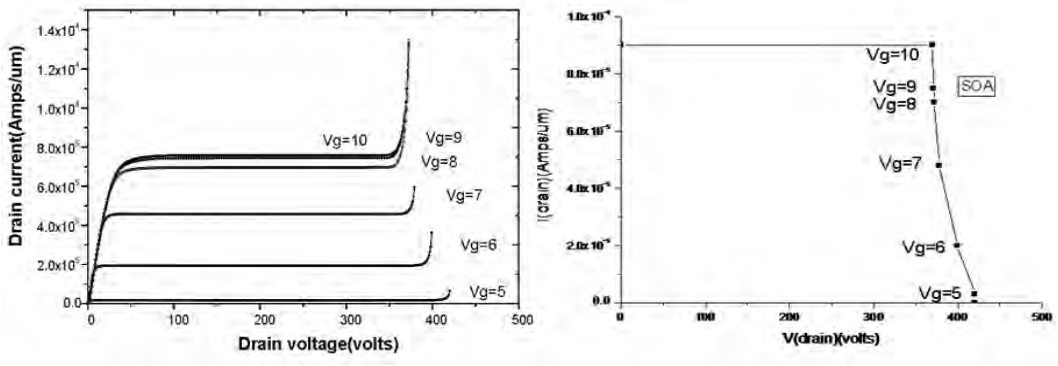


(a) (b)

Fig. 5 The dependence of optimal breakdown voltage (a) and the corresponding doping gradient slope G (b) on the drift region length.

Fig. 6(a) shows the drain current versus the drain voltage for different gate voltages, and Fig. 6(b) depicts the safe operation area (SOA) between the drain current and breakdown voltage for a device with an 60μm drift region.

Ex. 1016, Page 160



(a) (b)

Fig. 6 The drain current versus drain voltage at different gate voltage (a) and the safe operating area (SOA) of drain current versus on-state breakdown voltage (b).

Fig. 7 shows the figure of merit (FOM) between the on-state resistance and the breakdown voltage on a log-log scale for devices with drift lengths of 30, 40, 50, and 60um at 27℃ (a) and 125 ℃ (b). The α and β for 27℃ are 95,and 1.5 respectively;. The α and β for 125℃ are 187 and 1.5 respectively. The α and β are also compared with the previously reported data published [9,12]

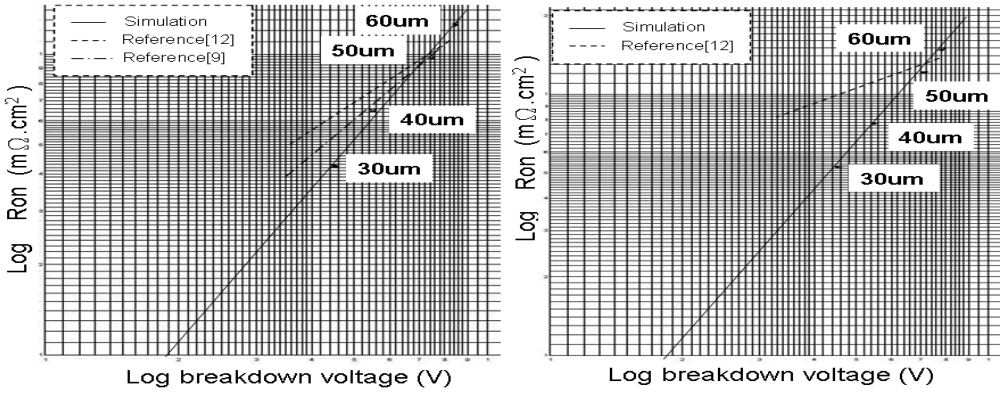


Fig. 7 Figure of merit (FOM) between the on-state resistance and the breakdown voltage on log-log scale for devices with drift lengths of 30, 40, 50, and 60um at 27℃(a) and 125℃(b).

Conclusion

The breakdown voltage dependence on doping density and concentration slope in the drift region with lengths of 30, 40, 50, and 60um are thoroughly simulated by TSuprem4 and Medici. Table 2 shows the summary of the optimal doping slope G or implant dose in the drift region to achieve the best on/off state breakdown voltage and the corresponding on-state resistance. In the future a test device needs to be designed to validate the simulated results.

Ex. 1016, Page 161

Drift length (μm)	G (cm^{-4})	Drift Dose (cm^{-2})	BV (off) (V)	BV (on) (V)	Vt (v)	Ron ($m\Omega.cm^2$)
30	4.20E+20	9.50E+12	459	237	5	41
40	2.90E+20	8.70E+12	519	146	5	63
50	2.40E+20	9.00E+12	743	560	5	98
60	2.10E+20	9.45E+12	810	596	5	127

Table 2 Summary of the optimal doping slope G or implant dose in the drift region to achieve the best on/off state breakdown voltage and on-state resistance.

References

1. Y. F. Guo, A new two dimensional analytical breakdown model of SOI RESURF devices, IEEE ICCCAS , pp. 1278-1282 (2007)
2. J. HE, Linearly graded doping drift region: a novel lateral voltage-sustaining layer used for improvement of RESURF LDMOS transistor performances, Semiconductor science and technology, vol. 17, pp. 721-728 (2002)
3. S. Merchant, Dependence of breakdown voltage on drift length and buried oxide thickness in SOI RESURF LDMOS transistors, ISPSD, pp. 124-128 (1993)
4. X. R. Luo, A novel analytical model for the potential and the electric field distribution of TFSOI power device, IEEE ICCCAS , pp. 1407- 1411 (2005)
5. S. D. Zhang, Numerical modeling of linear doping profiles for high-voltage thin-film SOI devices, IEEE Trans. Electron Dev., vol. 46, pp. 1036-1041 (1999)
6. HE Jin, A Novel Analytical Model for Surface Electrical Field Distribution, CHINESE JOURNAL OF SEMICONDUCTORS, vol.22,No.4, pp. 402-408 (2001)
7. M. Bawedin, LDMOS in SOI technology with very-thin silicon film . Solid-State Electronics , vol. 48, issue 12, pp. 2263-2270 (2004)
8. W. Yang, A novel analytical model for the breakdown voltage of thin-film SOI power MOSFETs. Solid-state electronics , Solid-State Electronics, pp. 43-48 (2005)
9. D. R. Disney, A new 800 V lateral MOSFET with dual conduction paths , IEEE ISPSD, pp. 399-402 (2002)
10. J. P. Colinge, Electron Lett., vol. 22, pp. 187-188 (1986)
11. F. Jian, Study for safe operating area of high voltage LDMOS, IEEE Business of Electronic Product Reliability and Liability, pp. 15- 18 (2004)
12. M.M Iqbal, F. Udrea, Technology-Based Figure of Merit (FOM) for High Voltage LDMOSFETs - Proof of Value of SOI in Power ICs , IEEE SOI Conference, pp. 57-58 (2007)
13. H. Shyam, Realizing high-voltage junction isolated LDMOS transistors with variation in lateral doping, IEEE Trans. Electron Dev., vol. 51, pp. 2223-2228 (2004)

Ex. 1016, Page 162

ECS Transactions, 18 (1) 135-140 (2009)
10.1149/1.3096441 © The Electrochemical Society

Investigation of Power MOSFET with Strained SiGe Channel

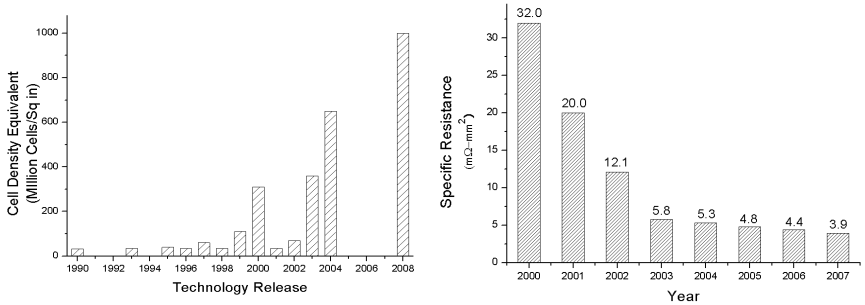
Shan Sun[a], Wei Zhou[b,c], Liren Yan[b,c], Jun Xu[b,c], Jiann S,Yuan[a], and Zheng J. Shen[a]

[a] University of Central Florida, Orlando, Florida, USA
[b] Tsinghua National Laboratory for Information Science and Technology
[c] Institute of Microelectronics, Tsinghua University, Beijing, China

Two trench-gated power MOSFETs using strained SiGe channel are proposed to further reduce specific on-state resistance. The first is a multiple SiGe and Si layer structure for N-channel MOSFET. A 30nm $Si_{0.85}Ge_{0.15}$ layer and a 50nm Silicon layer are grown by RPCVD alternatively to achieve sufficient channel length which is longer than the critical thickness before the trench etch. The second method is to form the strained SiGe layer along the trench sidewall for P-channel MOSFET. A single $Si_{0.85}Ge_{0.15}$ layer of 30nm thick is grown after the trench etch. It is expected that the channel mobility in the vertical direction will be increased by up to 40% and 60% for electrons and holes, respectively, compared to conventional silicon device with same structure and doping profile. The on-state resistance is predicted to be reduced by 12% and 20% for N-channel and P-channel MOSFET accordingly.

I. Introduction

Power semiconductors play a very important role in power management applications. Power MOSFETs are designed to provide low specific on-state resistance with certain blocking voltage. This trade-off is one of the most important factors for improving device performance. To reduce on-state resistance, optimization of structure, layout and process technology are widely under research in the past two decades [1-5]. One major method to achieve lower specific resistance is to increase cell density. Trench-gated powers MOSFETs with very high cell density dominate the low voltage power MOSFET market in recent years. The cell density has been increased and the specific resistance has been decreased steadily from 32mΩ-mm^2 to as low as 3.9mΩ-mm^2 as shown in Fig.1 [6-7]. However, this trend is considerably slowed down in recent years due to the theoretical limitation of silicon material.



(a) (b)

Fig.1 (a) Improvement in cell density of 30V n-channel MOSFET (b) Trend of on state resistance reduction

Ex. 1016, Page 163

Strained engineering is well known to enhance carrier mobility and is widely used in CMOS devices. Tensile strained silicon exhibits mobility enhancement for both carriers. Compressive strained SiGe provides significant in-plane hole mobility enhancement and out-of-plane electron mobility enhancement [8-10]. Strained engineering is therefore potentially able to further reduce the on-resistance of power MOSFETs. However, the application of strained MOS channel in power MOSFETs is limited by two mechanisms. First, the channel of a power MOSFET needs to be sufficiently long in order to support the needed blocking voltage, which unfortunately exceeds the SiGe relaxation critical thickness. Secondly, strained SiGe wafers require a thermal budget lower than conventional silicon technology. A solution of Ge implantation after high temperature diffusion processes [8] is adopted to overcome this issue. However, one challenge for this method is the control of Ge distribution and resulted strain performance.

This paper proposes an N-channel and P-channel trench-gated power MOSFET taking advantage of strained SiGe to reduce specific on-state resistance. Reduced Pressure Chemical Vapor Deposition (RPCVD) is used to grow stable strained SiGe layer. A novel low temperature SiGe process compatible with current silicon technology is also developed. Comprehensive device simulation and modeling are also carried out.

II. Device Structures

Compressively strained SiGe provides mobility enhancement for electrons in out-of-plane direction and holes in in-plane direction. Thus two different structures are proposed for N- and P- channel MOSFET respectively. A multi-layer structure is proposed for N-channel MOSFET in Fig.2 (a) features a channel with alternatively deposited Si and SiGe layer. In the repeated multi-layers, SiGe is compressively strained by Si layer in the lateral plane. The electron mobility is thus enhanced in the vertical out-of-plane direction, and the channel resistance is reduced accordingly. The other structure proposed, as in Fig.2 (b) for P-channel MOSFET features one single strained SiGe layer deposited after the trench is formed. The SiGe layer along the trench sidewall is under compressive strain by the Si trench in the vertical plane. The increased in-plane hole mobility leads to decreased channel resistance.

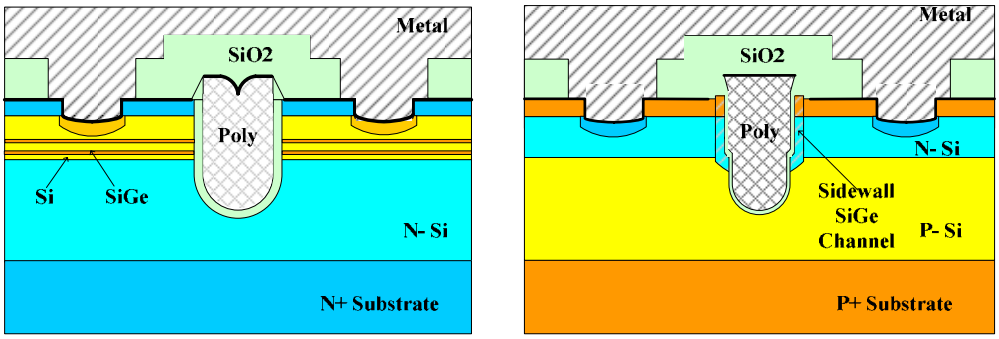


Fig.2 (a) Multi-layer N-channel MOSFET (b) Sidewall SiGe P-channel MOSFET

III. Fabrication Processes

A. Multilayer Structure for N-channel MOSFET

Fig.3 shows the process flow for the SiGe multi-layer N-channel power MOSFET. The starting material is 0.003Ω-cm {100} N+ Si substrate. An N-type epitaxial layer of 0.21Ω-cm (2e16/cm^3) is grown as the drift region. The Ge mole fraction is chosen to be 0.15 due to equipment limitation. Subsequently, the channel strained layers are deposited by RPCVD starts from a 10nm SiGe buffer layer, followed by 30nm $Si_{0.85}Ge_{0.15}$ and 50nm of Si with in-situ Boron doping of 2.5e17/cm^3. The deposition is performed two more times to achieve the 0.25um channel length required by blocking voltage. A 0.7um deep, 0.8um wide trench is then etched and the surface is smoothed by wet treatment. A pad thermal oxide is grown at 800ºC for 30 minutes followed by 250Å SiO_2 layer deposition as the gate oxide. Arsenic implantation at 70keV, 4e13/cm^2 and Boron implantation at 50keV, 2e14/cm^2 are used for source region formation. Rapid thermal anneal is employed at 900ºC for 30 seconds to prevent SiGe from relaxation.

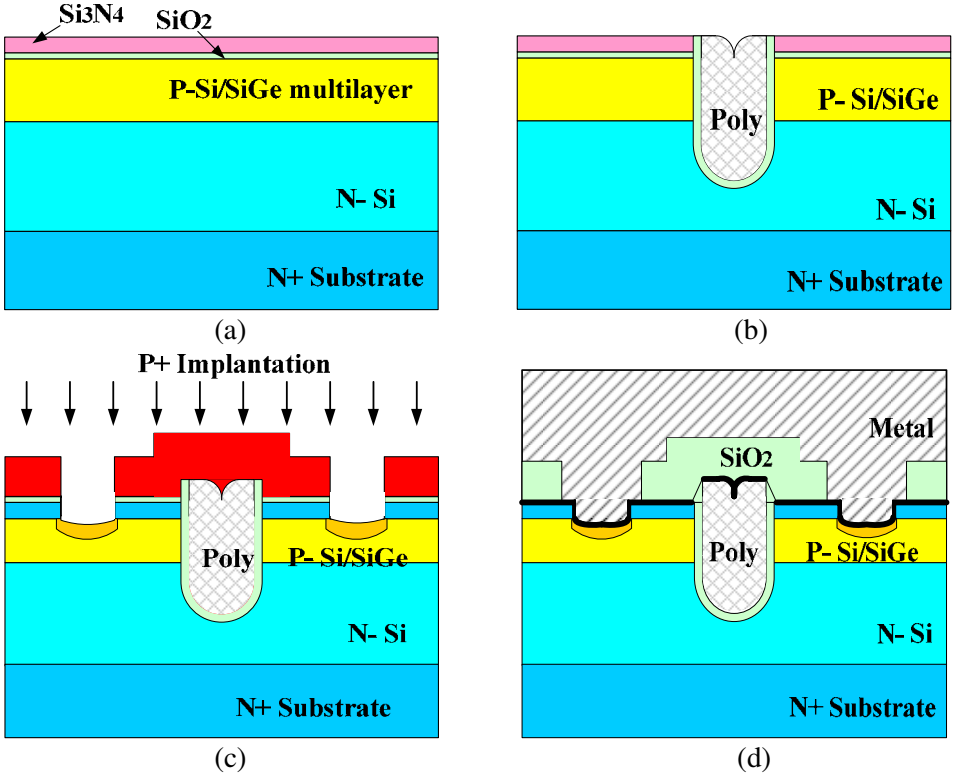


Fig.3 The process flow of Strained SiGe power MOSFET (a) multilayer formation (b) trench etch and poly fill (c) source recess etch and implantation (d) TEOS deposition, contact window etch and metal patterning.

Process simulation is carried out to achieve the proposed device structure followed by device simulation on the achieved device structure. Fig.4 illustrates the comparison of carrier mobility of SiGe and conventional Si device. The mobility of SiGe is enhanced by up to 40% in the channel region. The valleys of the strained device mobility are resulted

from disruption of the strain profile at SiGe and oxide interface. With the mobility enhancement, the specific on-state resistance of the strained SiGe device is reduced by 12% comparing to the silicon device from 13.7mΩ-mm^2 to 12.1mΩ-mm^2. The breakdown voltage and gate charge are unaffected and therefore no blocking and switching capability will be sacrificed.

The impact of mole fraction on device performance is also examined. Fig.5 shows that with Ge mole faction increasing from 0 to 0.2, the breakdown voltage and gate charge are not significantly changed while the specific resistance is further reduced. Therefore, a higher Ge concentration is preferred to provide further resistance reduction.

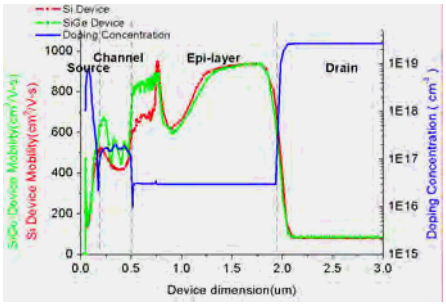


Fig.4 Mobility near the oxide/SiGe interface

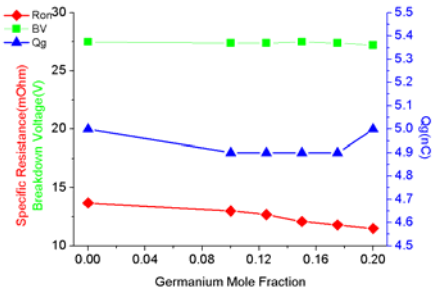


Fig.5 Impact of mole fraction

B. Sidewall SiGe Channel Structure for P-channel MOSFET

Fig.6 shows the process flow for trench MOSFET with sidewall SiGe channel. The starting material is 0.003Ω-cm {100} P+ Si substrate. A P-type epi-layer of 1.45 Ω-cm (1e16/cm^3) is grown as the drift region. A two-stage trench etch is implemented to form the SiGe channel. First, a 0.8 um wide trench of 0.7um deep is etched and the surface is smoothed by wet treatment to facilitate the growth of high quality SiGe layer which is achieved by selective epitaxial growth (SEG). A 10nm buffer layer, a 30nm strained $Si_{0.85}Ge_{0.15}$ of 30nm with in-situ phosphorus doping of 3e17/cm^3 are deposited. A Si_3N_4 is spacer is formed on the sidewall of the trench and a second etch is performed to remove SiGe from the bottom. The trench is wet treated again followed by a thin Si cap layer to grown high quality oxide. Subsequently, low temperature oxide is grown at 800ºC for 30min, and 250 Å SiO_2 is deposited as the gate oxide. Boron implantation at 40keV, 4e13/ cm^2 is performed, a recess is opened at the implanted area and Arsenic implantation at 70keV, 4e13/cm^2 is performed to form the source region which helps prevent latch up event. A rapid thermal anneal process is employed at 900ºC for 30 seconds.

The proposed structure is achieved by process simulation and device performance is inspected. Fig.7 shows the mobility comparison for SiGe and Si devices. The hole mobility is improved by as much as 60% in the channel region. The mobility improvement is achieved along the whole channel region without disruption of strain at the SiGe and oxide interface. The specific on-state resistance is decreased significantly by 20%, from 27.07mΩ-mm^2 for silicon to 21.6mΩ-mm^2 while breakdown voltage reduced slightly by 4% from -32.5V to -31.2V. The gate charge exhibits minor change and the switching performance remains unaffected.

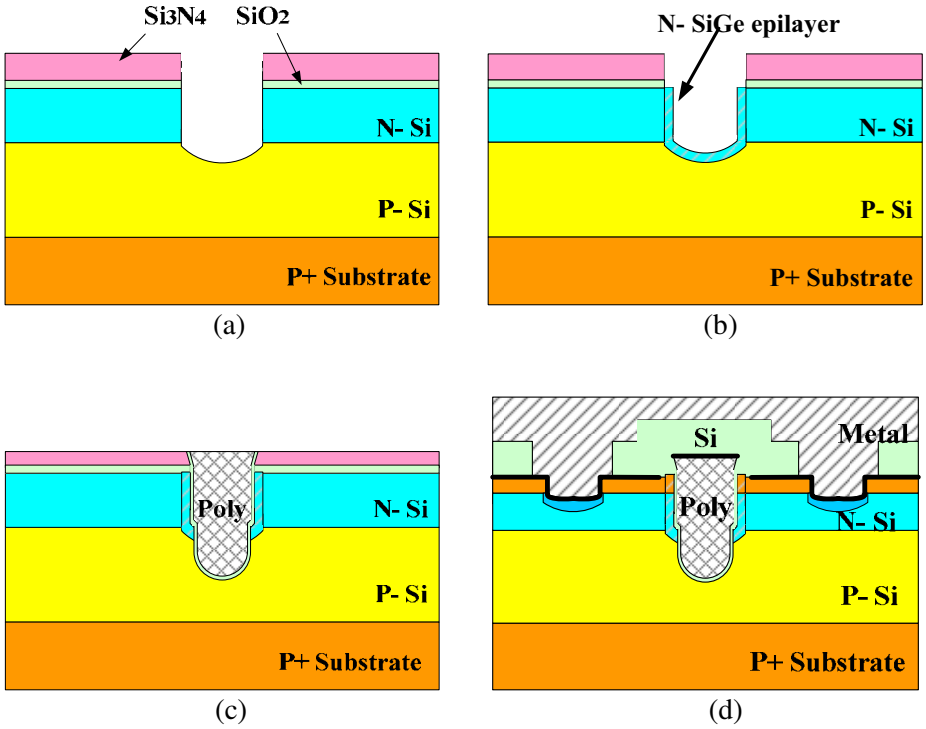


Fig.6 Process flow of SiGe sidewall channel trench MOSFET (a) trench etch (b) N-SiGe epi-layer deposition (c) second-step trench etch and poly fill (d) source implantation and TEOS, metal formation

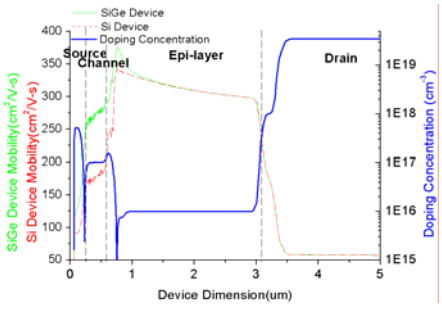


Fig.7 Mobility near the oxide/SiGe interface

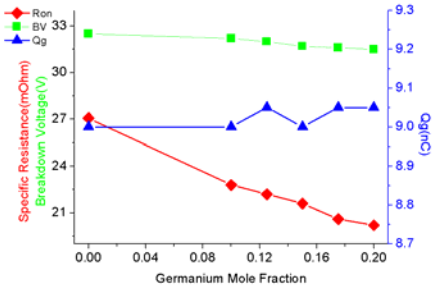


Fig.8 Impact of Ge mole fraction

The Ge mole fraction impact is shown in Fig.8. With the increasing Ge mole fraction, the specific resistance decreases rapidly while breakdown voltage and gate charge variation are observed to be insignificant.

IV. Summary

Strain engineering is applied to trench power MOSFETs in this paper. Two new device structures are proposed to implement strained SiGe on both N-channel and P-

Ex. 1016, Page 167

channel devices. Comprehensive simulation and modeling was carried out to predict device performance with strained SiGe channel. Specific on-state resistance is reduced by 12% and 20% respectively for N-channel and P-channel MOSFET at a mole fraction of 0.15. The specific resistance is predicted to be further reduced by increasing the Ge concentration. A low temperature SiGe process which is compatible to current silicon technology is also developed in this work.

V. Acknowledgments

The authors wish to acknowledge the support from the National Science Foundation of China under Award 60636010, 60820106001, and from the Tsinghua National Laboratory for Information Science and Technology and Institute of Microelectronics at Tsinghua University for process development. This work was also supported in part by the U.S. National Science Foundation under Award ECS-0454835 and a grant from Intel Corporation.

VI. References

1. Jun Zheng, 13thInternational Symposium on Power Semiconductor Devices and IC's, 147, (2001)
2. Naoto Fujishima, IEEE transactions on Electron Devices, vol.49, no.8, 1462 (2002)
3. I-ming Lee, J. Electrochemical. Soc., vol.144, no. 3, 1095 (1997)
4. Jinggang Lu, Electrochemical Society Transactions, vol.11, no.3, 29 (2007)
5. P.Goarin, 19th International Symposium on Power Semiconductor Devices and IC's, 61 (2007)
6. D. Kinzer, Microelectronics Journal, vol.3, no.225, 35 (2004)
7. G.MOXEY, http://www2.electronicproducts.com/Dc_dc_converter_efficiency_revisited-article-FAJH_Energy_Jul2008-html.aspx
8. Xiangdong Chen, IEEE transactions on Electron Devices, vol.48, no.9, 1975 (2001)
9. M.H.Juang, Semiconductor Science and Technology, vol.21, no.6, 799 (2006)
10. Yasuhiro Shiraki, Surface Science Reports, vol.59, issues 7-8, 153 (2005)

Ex. 1016, Page 168

ECS Transactions, 18 (1) 141-148 (2009)
10.1149/1.3096442 © The Electrochemical Society

Simulation of retention characteristics for metal nanocrystal nonvolatile memory with a modified direct tunneling model

Yingtao Li[a, b], Ming Liu[a,*], Shibing Long[a], Qin Wang[a], Qi Liu[a], Sen Zhang[a], Yan Wang[a, b], Qingyun Zuo[a], Su Liu[b]

[a] Laboratory of Nano-fabrication and Novel Devices Integrated Technology, Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China
[b] Institute of microelectronics, School of physical science and technology, Lanzhou University, Lanzhou 730000, China

By comparing the impacts of Coulomb blockade effect and quantum confinement effect on the energy band shift of metal nanocrystal, it is obvious that not only quantum confinement effect but Coulomb blockade effect could influence the energy band shift of metal nanocrystal. Taking into consideration these two effects, a modified direct tunneling model and numerical calculations is proposed for describing the retention characteristics of metal nanocrystal nonvolatile memories completely. The simulation results which compare well with experimental data show that the retention characteristics are not only dependent on tunneling dielectric thickness, but are also related to the barrier height, dielectric constant, as well as the nanocrystal coverage ratio.

Introduction

Since Kahng and Sze proposed the first floating gate semiconductor nonvolatile memory (NVM) device in 1967 (1), the research attention of NVM devices has been considered as one of the most important things from the end of 20th century to now. Along with the development of nonvolatile memory (NVM) into the nano-scale filed, nanocrystals (NC) such as semiconductor (2-6), metal (7-10) or hetero-nanocrystal (11, 12) used as a floating gate are being enthusiastically investigated in recent years. Due to the nanocrystal approach consisting in replacing the continuous floating gate used in the conventional Flash memory cell by discrete charge storage nodes, NC memory devices are virtually immune from anomalous charge loss, and thus, they have several advantages, such as low operation voltage, better endurance, better retention performance, as well as faster writing and erasing speeds (8-16). It is well known that the retention time for a nonvolatile memory device is one of the most important parameters, and lots of researches have shown that metal NC-based NVM decives have a longer retention time. However, for a metal NC nonvolatile memory device, when the size of the nanocrystals comes into the range of a few nanometers, the effect of quantum confinement effect (17, 18) and Coulomb blockade effect (19) on the retention time can not be neglected.

In this letter, we present a new numerical model which is extended to include both the quantum confinement effect and the Coulomb blockade effect. Otherwise, the retention characteristics of metal-NC-based NVM devices are simulated using this model. And the simulation results correctly account for the retention characteristics dependence of work

Ex. 1016, Page 169

function of NCs, different dielectric materials, as well as the nanocrystal coverage ratio.

Theory and model

Quantum confinement effect and Coulomb blockade effect

With the size of the metal nanocrystals coming into the nanometer range, the quantum confinement effect and Coulomb blockade effect in the energy band structure of the metal nanocrystals becomes a significant phenomenon. For quantum confinement effect, which can cause the metal nanocrystal energy band edge shift to higher energy compared with bulk material, and as a result, the energy band offset between the nanocrystal and the surrounding dielectric is reduced. For Metal-NC, according to carefully researched electronic properties of metallic fine particles by Kubo (20), the quantum confinement effect makes the continuous energy band of Metal-NC split into separated energy levels, and the average spacing δ can be given by Eq. [1]：

$$\delta = \frac{4E_F}{3N} \qquad [1]$$

where E_F is the Fermi level of the bulk metal, N which can be calculated by Eq. [2] is the total number of conductive electrons in the nanocrystal.

$$N=\int_0^{E_F} N(E)dE \qquad [2]$$

$$N(E) = 4\pi V \frac{(2m^*)^{3/2}}{h^3} E^{1/2} \qquad [3]$$

where N(E) is the density-of-states of conductive electrons， V is the volume of metal-NC, m^* is the effective mass of electron in the matal-NC, h is the Planck's constant.

Assuming m* equal to the mass of a free electron and the metal-NC is spherical, substituting Eq. [2] and [3] in Eq. [1], we obtain that the energy band shift induced by quantum confinement effect is:

$$\delta = \frac{0.5625}{d^3\sqrt{E_F}}(eV) \qquad [4]$$

On the other hand, the Coulomb blockade effect can effectively inhibit the electron tunneling. When some electrons tunneling across the tunneling dielectric layer and stored in nanocrystals, the electron charge will raise the nanocrystal energy level. This raising of energy level will reduce the electrical field across the tunneling dielectric layer and prevent others electrons tunneling across the tunneling dielectric layer (19). Otherwise, if the nanocrystal potential energy is high enough, the electrons stored in the nanocrystals have large tendency to tunnel back into the channel during retention, leading to a weak retention characteristic. According to the result carefully researched by Steimle et al. (19), the energy level change $\Delta\mu$ resulting from Coulomb blockade effect is given by:

$$\Delta\mu = \frac{q^2}{C} \qquad [5]$$

where $C = 2\pi\varepsilon_{tun}d$ (18) is the self capacitance of NC, ε_{tun} is the dielectric constant of tunneling dielectric and d is the diameter of NC.

The effects of Coulomb blockade effect (CBE) and quantum confinement effect (QCE) on the Fermi level shift of various Metal-NCs are shown in Figure 1, respectively. For

the same $Si0_2$ as tunneling dielectrics, when the diameter of NC is much smaller than 0.6 nm, QCE is dominant, but with the diameter increased, the shift of the energy band caused by CBE is more prominent shown in Figure1 (a). In addition, the effects of various tunneling dielectrics on the energy band shift resulting from Coulomb blockade effect are shown in Figure 1(b). It should be noticed that the lower the dielectric constant of tunneling dielectric is, the more serious the conduction band shift. The dielectric constants of various oxide tunneling dielectrics are shown in Table 1.

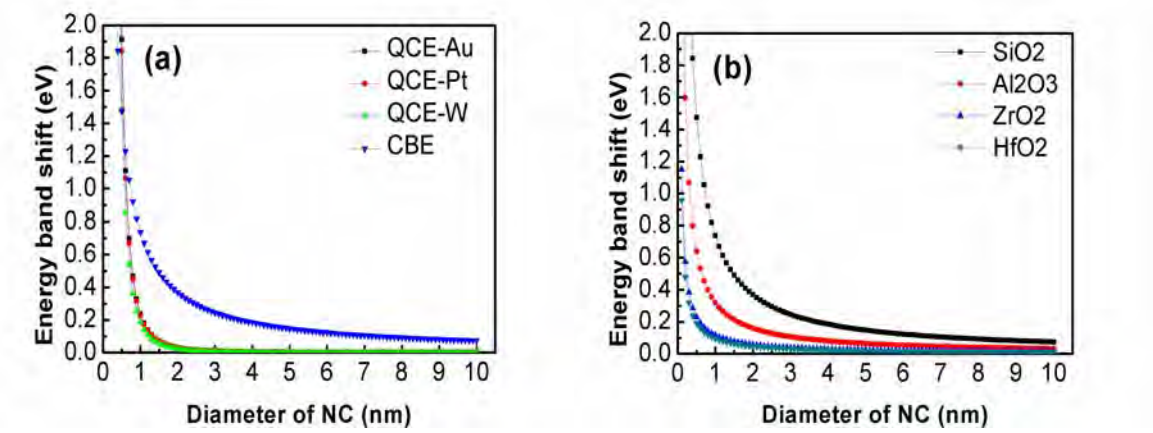


Figure 1 The energy level shift of Metal-NCs. (a) caused by CBE and QCE with the same $Si0_2$ tunneling dielectric, and (b) caused by CBE with various tunneling dielectrics.

Table 1. The relative permittivity (ε) of tunneling dielectric, and the barrier height for electrons (ϕ^e) at Metal-NC/tunneling dielectric interface.

Material	ε	W-ϕ^e (eV)	Au-ϕ^e(eV)	Pt-ϕ^e (eV)
SiO_2	3.9	3.5	4.1	4.65
Al_2O_3	9	2.6	3.2	3.75
ZrO_2	25	1.7	2.3	2.85
HfO_2	30	1.8	2.4	2.95

Direct tunneling model

In our model, owing to the control gate is zero biased during retention and the number of charges stored in the nanocrystals is limited, the electric field in the tunneling dielectric is low. Furthermore, with the scaling of NC nonvolatile memory, tunnel oxide for nonvolatile memory has to be scaled. And thus, the direct tunneling (DT) effect becomes dominant (21). In the derivation of the DT theory, the standard expression for current density is (22):

$$J_{DT} = AE_{tun}^2 \exp\left\{-\frac{B\phi_b^{3/2}}{E_{tun}}\left[1-\left(1-\frac{qV_{tun}}{\phi_b}\right)^{3/2}\right]\right\} \quad [6]$$

where $A = \frac{q^3 m_0}{8\pi h m^*_{tun} \phi_b}$ and $B = \frac{8\pi\sqrt{2m^*_{tun}}}{3hq}$, m_0 is the mass of a free electron, m^*_{tun} is the effective mass of electron in tunneling dielectric, and ϕ_b is the barrier height.

According to the results analyzed above, a theoretical numerical model using direct tunneling theory with both CBE and QCE is proposed for discussing the retention characteristics of Metal-NC nonvolatile memory. When taking into consideration the impact of CBE and QCE, the current density expression Eq. [6] for DT is no longer valid but becomes:

$$J_{DT} = AE^2_{tun} \exp\left\{-\frac{B\phi_b^{*3/2}}{E_{tun}}\left[1-\left(1-\frac{qV_{tun}}{\phi_b^*}\right)^{3/2}\right]\right\} \quad [7]$$

where $\phi_b^* = \phi_b - (\Delta\mu + \delta)$, V_{tun} is the voltage drop across tunneling dielectric layer which can be written as $V_{tun} = E_{tun} t_{tun}$, and E_{tun} which can be calculated by Eq. [8] is the electrical field in tunneling dielectric t_{tun} is the thickness of tunneling dielectric layer.

$$E_{tun} = \frac{Q}{\varepsilon_{tun} R} \quad [8]$$

where ε_{tun} is the dielectric constant of tunneling dielectric and R is the nanocrystal coverage ratio. With electrons tunneling out of the nanocrystals, the charge density Q decreases with the time, and the change between Q and J_{DT} can be calculated by:

$$\frac{dQ}{dt} = -J_{DT} \quad [9]$$

Otherwise, the threshold voltage shift due to electron storage and tunneling is given by (2)

$$\Delta V_t = \frac{Q}{\varepsilon_{tun}}\left(t_{con} + \frac{\varepsilon_{tun} d}{2\varepsilon_{nc}}\right) \quad [10]$$

Using the modified model, and associating with the correlative equations together, we can numerically calculate the threshold voltage shift ΔV_t as well as the relationship between ΔV_t and time, and thus the retention time of nanocrystal memory.

Results and discussion

In this paper, the retention time is defined as the time when 50% of the charge leaks at zero gate bias. And, the parameters used in simulation were as follows: the control dielectric thickness t_{con} = 20 nm, the initial value of the threshold voltage $\Delta V_t(0)$ = 1 V, V_{tun} = 4 nm, R=50% and Metal-nanocrystal diameter d = 4 nm.

Figure 2 shows the experimental and simulated threshold voltage shift (ΔV_t) as a function of time. The experimental data used in the simulation is extracted from the work of Lee (23) and Takata et al. (24), respectively. As it is shown in Figure 2, it is clear that the simulation results with Coulomb blockade effect are in much better agreement with experimental data compared to the results without Coulomb blockade effect, especially in Figure 2(b), which demonstrates that the modified numerical model is correct for calculating the retention characteristics of Metal-NC memory devices.

Ex. 1016, Page 172

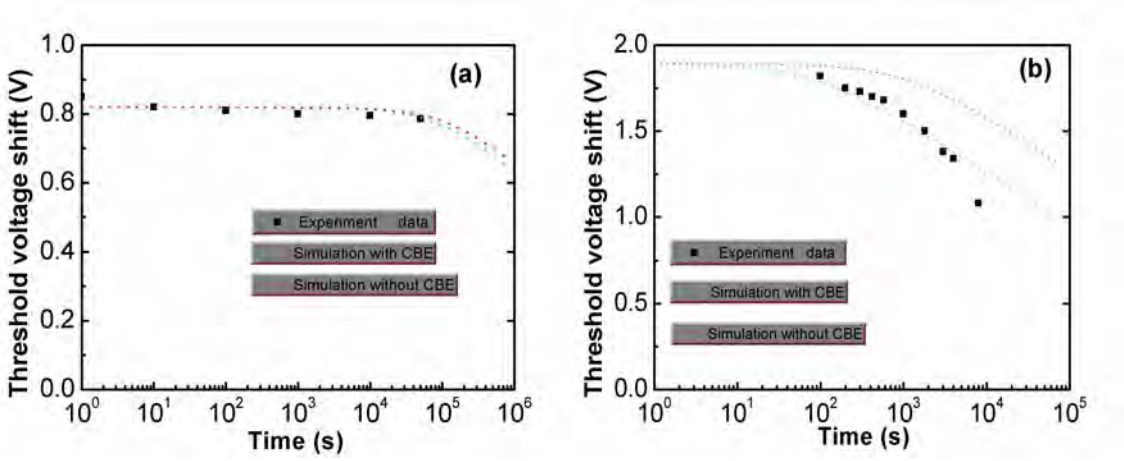


Figure 2 Experimental data and simulated results with CBE and without CBE. (a) the experimental data was from Lee (23), and (b) from Takata et al (24).

Figure 3 (a) shows the curve of retention time versus various Metal-nanocrystals (W, Au and Pt) and tunneling dielectric materials, respectively. For Metal-NC memory with the same tunneling dielectric, the simulation results show that Pt-NC has a better retention characteristic than that of Au-NC and W-NC. This result mainly attributes to the larger work function of Pt material and thus much higher barrier height for electrons at Pt/tunneling dielectric interface, which is shown in Table 1. Due to the higher barrier height, a deeper potential well is current in the Pt-NC device, which directly results in the fact that electrons captured by the Pt-NC are more difficult to discharge from the floating gate. As a result, there is lower leakage current in Pt-NC device and thus a longer retention time.

Otherwise, taking Au-NC as an example to illustrate the impact of various tunneling dielectrics with the same equivalent oxide thickness (EOT) on the retention characteristic is shown in Figure 3 (b). When using the same EOT (equal to 1nm), the retention characteristic of the Au-NC device which is of high permittivity material (HfO_2) as tunneling dielectric is much better and more prominent compared with the low permittivity devices. There are three reasons contributing to this result: First, a low electric field is formed in the high-k tunneling dielectric layer during retention and thus weakening the leakage current. The second reason is the weaker Coulomb blockade effect for the high-k dielectric as it is shown in Figure 1(b). Another more important reason is the thicker physical thickness of HfO_2 when utilizing the same EOT, and thus improve the device retention characteristic.

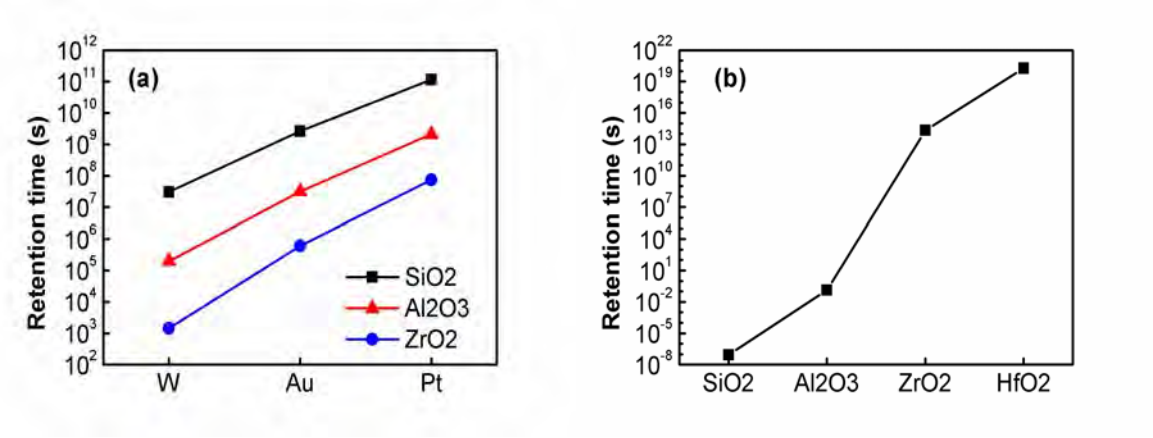


Figure 3 The curve of retention time versus various Metal-nanocrystals and tunneling dielectric materials (a), and various tunneling dielectric with the same EOT (b).

The impact of nanocrystal coverage ratio on the retention characteristic is shown in Figure 4. For a dedicated tunneling dielectric material (SiO_2), when all the other parameters are the same, it is noted that the retention time of the Metal-NC memory increases with the increase of the nanocrystal coverage ratio. That is because the large nanocrystal coverage ratio means more charge retention sites existing in the floating gate, and thus much more electrons can be captured and stored in the NCs during retention. Therefore, a lower leakage current and a better retention characteristic are obtained for a larger nanocrystal coverage ratio.

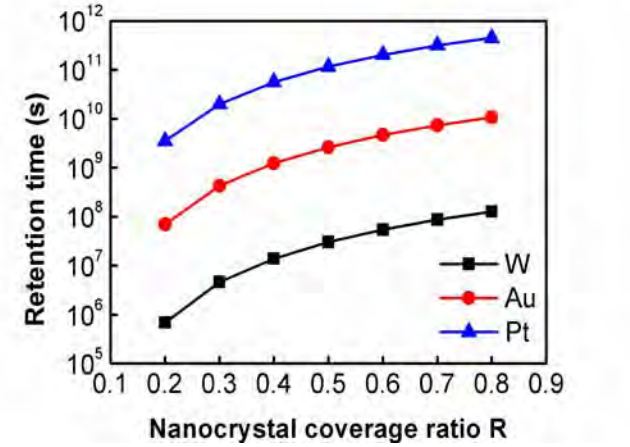


Figure 4 The impact of nanocrystal coverage ratio on the retention characteristic.

Conclusion

Taking into consideration the impacts of Coulomb blockade effect and quantum confinement effect on the energy band shift of a metal nanocrystal, a numerical calculations model including both these two effects was presented based on direct

tunneling theory. The calculated results compare well with experimental data. Corresponding to the modified model, the retention characteristics of Metal-nanocrystal nonvolatile memory devices can be completed. Owing to the less energy band shift caused by the Coulomb blockade effect for the high permittivity material, utilizing high-k materials as tunneling dielectric has a longer retention time for the same EOT. Otherwise, the model also accounts for the retention characteristics of nanocrystal nonvolatile memory devices prominent dependence of the tunneling dielectric thickness, barrier height, dielectric constant, as well as the nanocrystal coverage ratio.

Acknowledgments

This work was supported by the Hi-Tech Research and Development Program of China (863 Program) under Grant No 2008AA031403, the National Basic Research Program of China (973 Program) under Grant No 2006CB302706 and the National Natural Science Foundation of China under Grant Nos 60825403, 90607022 and 60506005.

References

1. K. Kahng and S. M. Sze, IEEE Trans. Electron Devices, **14**, 629 (1967).
2. S. Tiwari, F. Rana, H. Hanafi, A. Hartsten, E. F. Crabbe, and K. Chan, Appl. Phys. Lett., **68**, 1377 (1996).
3. Y. King, T. King, and C. Hu, IEEE Electron Device Lett., **48**, 696 (2001).
4. K. Han, I. Kim, and H. Shin, IEEE Trans. Electron Devices, **48**, 874~879 (2001).
5. M. Kanoun, A. Souifi, T. Baron, and F. Mazen, Appl. Phys. Lett., **84**, 5079 (2004).
6. S. Choi, H. Yang, M. Chang, S. Baek, H. Hwang, S. Jeon, J. Kim, and C. Kim, Appl. Phys. Lett., **86,** 251901 (2005).
7. Z. Liu, C. Lee, V. Narayanan, G. Pei, and E. C. Kan, IEEE Trans. Electron Devices, **49,** 1606~1613 (2002).
8. C. C. Wang, J. Y. Wu, Y. K.Chiou, C. H. Chang, and T. B. Wu, Appl. Phys. Lett., **91,** 202110 (2007).
9. D. B. Farmer and R. G. Gordon, J. Appl. Phys., **101,** 124503 (2007).
10. J. Y. Tseng, C. W. Cheng, S. Y. Wang, and T. B. Wu, Appl. Phys. Lett., **85,** 2595 (2004).
11. D. Zhao, Y. Zhu, R. Li, and J. Liu, Solid-State Electronics, **50,** 362~366 (2006).
12. D. Zhao, Y. Zhu, R. Li, and J. Liu, Solid-State Electronics, **49,** 1974~1977 (2005).
13. J. J. Lee, X. G. Wang, W. P. Bai, N. Lu, and D. L. Kwong, IEEE Trans. Electron Devices, **50,** 2067~2072 (2003).
14. J. J. Lee and D. L. Kwong, IEEE Trans. Electron Devices, **52,** 507 (2005).
15. P. Punchaipetch, Y. Uraoka, and T. Fuyuki, Appl. Phys. Lett, **89,** 093502 (2006).
16. T. H. Ng, V. Ho, L. W. Teo, M. S. Tay, B. H. Koh, W. K. Chim, W. K. Choi, A. Y. Du, C. H. Tung, Thin solid films, **462-463**, 46~50 (2004).
17. Y. M. Niquet, G. Allan, C. Delerue, M. Lannoo, Appl Phys Lett., **77,** 1182~1184 (2000).

Ex. 1016, Page 175

18. T. Buuren, L. N. Dinh, L. L. Chase, W. J. Siekhaus, L. J. Terminello, Phys Rev Lett., **80**, 3803~3806 (1998).
19. R. F. Steimle, R. Muralidhar, R. Rao, M. Sadd, C. T. Swift, J. Yater, B. Hradsky, S. Straub, H. Gasquet, L. Vishnubhotla, E. J. Prinz, T. Merchant, B. Acred, K. Chang. B. E. White, Microelectronics Reliability, **47,** 585~592 (2007).
20. R. Kubo, Journal of the physical society of Japan, **17,** 975~986 (1962).
21. J. Jung and W. J. Cho, Semicond. Sci. Technol., **8,** 32~37 (2008).
22. K. F. Schuegraf, C. M. Hu, IEEE Trans. Electron Devices, **41,** 761~767 (1994).
23. J. J. Lee, D. L. Kwong. IEEE Trans. Electron Devices. **52,** 507~511 (2005).
24. M. Takata, S. Kondoh, T. Sakaguchi, H. Choi, J-C. Shim, H. Kurino and M. Koyanagi, IEEE IEDM Technical Digest, 22.5.1~22.5.4 (2003).

CHAPTER 2

METROLOGY, RELIABILITY AND TESTING

ECS Transactions, 18 (1) 151-160 (2009)
10.1149/1.3096443 © The Electrochemical Society

Metrology Requirements for the 32 nm Technology Node and Beyond

John Allgair[a], Benjamin Bunday[b], Aaron Cordes[b], Pete Lipscomb[b], Milt Godwin[b], Victor Vartanian[b], Michael Bishop[b], Doron Arazi[c], Kye-Weon Kim[d]

[a] International SEMATECH Manufacturing Initiative (ISMI)/AMD assignee
[b]ISMI
[c]ISMI/Spansion
[d]ISMI/Samsung

Successful in-line metrology is imperative for a fab to achieve profitable production yields. Full functionality and high circuit speed are achieved only through control of defectivity and tight distributions of feature sizes. In-line monitoring of applicable metrics is key to ensuring success and is also used to fine-tune production processes for improved yield and circuit speed. Metrology has now become an inherent part of mission-critical production processes. This article gives a high-level overview of the findings of the ISMI metrology program and details some of the major approaching manufacturing challenges.

INTRODUCTION

Integrated device manufacturers will face a host of metrology challenges as manufacturing moves to the 32 nm and beyond. Some of these challenges stem from the usual shrinkage of feature sizes and design rules, but are complicated by the fact that many standard workhorse technologies are being pushed to their very limits. However, many new techniques, such as immersion lithography, optical proximity correction (OPC), design-based metrology (DBM), double-patterning (DP), strained channel devices, multi-gate FinFETs, and others, are introducing their own unique metrology requirements.

The International SEMATECH Manufacturing Initiative (ISMI) has the goal of developing a world class industry collaborative program for the development and application of measurement methods for advanced semiconductor technologies. Metrology projects underway at ISMI are aimed at assessing and improving the manufacturing worthiness of in-line measurement equipment. We accomplish this through facilitating working groups of industry experts, such as the Advanced Litho Metrology Advisory Group (AMAG) and its sister groups FMAG and DMAG (for thin film and defect metrology, respectively). These lead to the creation of industry-consensus unified specifications, leading-edge metrology test artifacts, rigorous and challenging tool evaluations, and extendibility studies and modeling. From these activities, we derive best known methods (BKMs), metrics, standards, and models for metrology performance, which enable us to make recommendations for tool development and improvement [1, 2]. ISMI membership represents a constantly growing selection of companies from across the globe, who collectively represent a solid majority of the world's integrated device manufacturing base.

This article reviews some of the major manufacturing challenges facing in-line semiconductor metrology at the 32 nm technology node and beyond. We identify

Ex. 1016, Page 179

challenges introduced by both shrinking design rules and new design elements and discuss future directions for development.

GENERAL METROLOGY ISSUES

Until recently, the most important metric by which many metrology tools were judged was precision—the "error bar" of a single measurement from a single tool. Generally speaking, tool precisions today meet the ITRS requirements. However, matching and accuracy are starting to approach precision in terms of real manufacturing importance.

Matching is the error bar of measurements performed on a fleet of tools. Ideally, all tools in a manufacturing bay need to be statistically indistinguishable. In practice, honest, real-life matching over the long term often shows fleet-wide precision to tolerance (P/T) ratios of 40%–50% or more.

Accuracy is the measure of how close a tool under test (TuT) agrees with a reference measurement system (RMS), which ideally is calibrated to be SI-traceable. Accuracy shows how well the TuT responds to real process change and thus how metrology tracks physical reality. This is becoming increasingly important for several reasons. First, OPC setup is a large driver of CD metrology; it is based on physical models, which in turn depend on measurements accurate in real-world units. Advanced process control (APC) also requires accurate reference measurements, as does optical CD (OCD, or scatterometer) model validation.

Meanwhile, throughput requirements are also going up. Increasing the statistical validity of accurate metrology results requires increased sampling. Move-acquire-measure (MAM) times of 1 second and less are becoming highly desirable.

450 mm Metrology Metrology of 450 mm wafers will be the same as 300mm wafers, since wafer size does not impact detection technology. The major technology development need will be wafer stages. To maintain or increase throughput, stages will need to be faster to account for larger distances. High precision stages are expected to maintain or enhance throughput for most metrology equipment.

Additional challenges for 450 will be related primarily to process node and starting materials. The 2007 ITRS identifies some potential challenges for starting materials metrology development: 1. Bulk Ni and Cu measurement on p+, silicon on insulator (SOI), strained silicon (SSi), and strained silicon on insulator (SSOI) wafers. 2. Measurement of 109–1010 cm-3 Fe (and other bulk metals) in the top Si of thin SOI wafers. 3. Thickness and uniformity of very thin SOI layers (<20 nm). 4. Defectivity of thin layers (e.g., threading dislocations, "HF defects"). 5. Particle detection (<100 nm) on layered surfaces.

Near-term tool development priorities to enable starting materials include 1. Edge inspection/review. 2. Macro inspection and optical review 3. Geometry measurement system (GBIR, Bow, Warp, TTV_9, Wafer Shape). 4. Defect inspection/detection (unpatterned wafers).

Ex. 1016, Page 180

To support development of 450 mm equipment, the ISMI Starting Materials project has created a 450 mm Wafer Bank. The 450 mm Starting Materials project currently has 100 mechanical grade sintered poly silicon wafers available and will lend wafers to supplier development programs.

LITHO METROLOGY

CD Metrology ISMI evaluation results show that, while CD single tool precision results have generally kept pace with the ITRS, matching and accuracy have not. Recently, based upon these findings, the ITRS metrics have changed from the traditional precision to uncertainty, which includes accuracy, matching, and sampling components [3].

Meanwhile, the measurement of simple linewidths is not sufficient anymore. See Figure 1. Many different measurements are of interest for process monitoring and OPC verification. OPC features lead to changes in shape such as different curvatures, line end pullbacks, and reaction of linewidth and pitch with topography. This is leading to a fundamental shift in SEM metrology, the metrology of contours, where shapes are compared to the intended design. Line edge roughness (LER), line width roughness (LWR), and line width variation (LWV) are also process variations of importance. Profile (in 3-D) is crucial for APC feedback.

Future device types will also raise new issues. FinFETs require true 3D metrology, as some of the critical features are related to the widths of metal and high-k dielectric layers partway up the sidewall of a fin. This can be addressed with OCD or to some extent by CD-SEM using pre-/post-measurement strategies. Etch recess under the fin (or under future gates in other device types) appears to be, to some extent, measurable by OCD, although shallow recesses might present low signal-to-noise issues. Metal gate devices may present challenges for OCD at some process steps due to the metal being opaque to ultraviolet light. Deep contact holes (with 6:1 aspect ratios and 30-40 nm diameters) also are becoming a larger issue, for both CD-SEM and OCD.

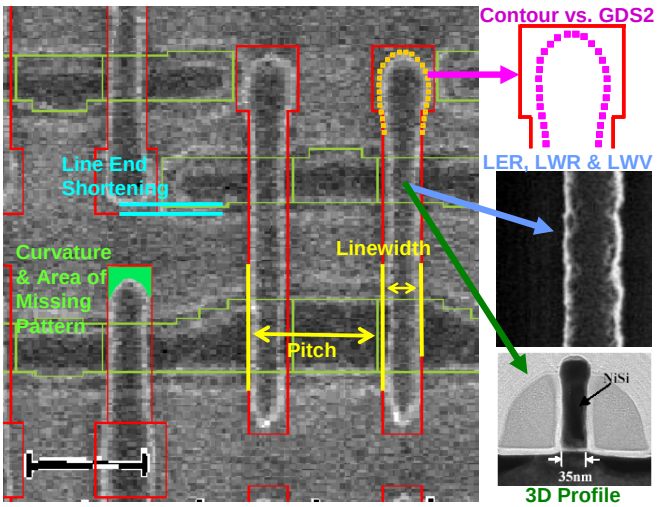


Figure 1: CD metrology is no longer limited to simple linewidths.

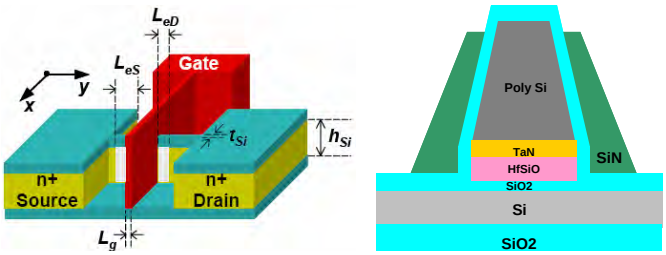


Figure 2: Advanced device architectures. (Left) FinFET. (Right) Metal gate.

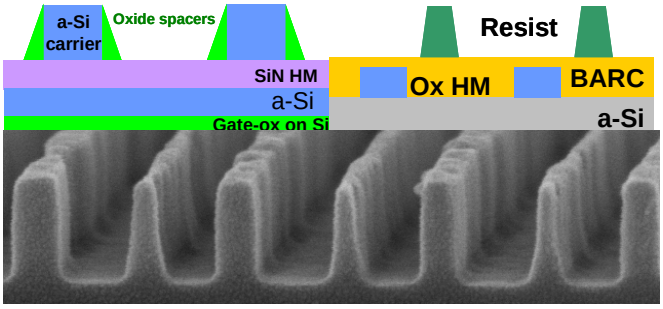


Figure 3: (Left) Spacer self-aligned DP, before removal of the sacrificial carriers. (Right) Dual litho/dual etch DP, at the natural process control step before final etch. (Bottom) DP structure after etch, showing dual populations of different parameters: CD, pitch, sidewall angle (SWA), LER, and height.

Another challenge to metrology is the advent of the common use of double patterning (DP). The spacer self-aligned type DP represents fewer challenges, although such etch-intensive processes, with some steps including partial etches, sometimes exhibit difficulty for building predictive APC models from such measurements. The dual litho/dual etch type DP is more challenging from a metrology perspective; at the natural process control step after the second litho exposure, alternate lines are of different optical properties (n and k) and separated by BARC, prohibiting CD-SEM from measuring both features at once for overlay. Many floating parameters will be necessary for OCD, as the features will exhibit two distinct populations of several key parameters such as CD, pitch, material, sidewall angle, roughness, and height. It is not yet clear if OCD will be necessarily sensitive to all these parameters. See Figure 3 for an illustration.

In general, 32 nm node toolsets will remain similar to the past—OCD will continue to have the key roll in process control, since it is a technique that gives good information about both average process CD and profile, and CD-SEM will also continue to give data on process variation and top-down feature shape in-circuit and be heavily used, in contour mode, to develop and monitor OPC performance.

Reference Metrology With increased accuracy comes an increased demand for reference metrology. Accurate reference metrology is needed to verify the accuracy of tools and to compare performance to maintain fleet matching standards. This is typically accomplished by using the reference tool to create calibration standards, which are then used to verify the performance of faster in-line metrology tools, such as CD-SEMs.

NIST and ISMI have jointly developed a CD-AFM based RMS that provides rigorously SI-traceable reference measurements for linewidth, height, pitch, and SWA [4].

Ex. 1016, Page 182

Overlay Metrology Traditionally, overlay metrology systems have taken their measurements from box-in-box targets located outside the circuit. However, box-in-box targets are inherently unstable. Upgrading to periodic grating-style structures has improved the precision of modern overlay systems into the 0.2 nm range, with sub-1 nm fleet uncertainty. This satisfies the ITRS overlay precision requirements of 0.50 nm for the 32 nm node.

The advent of DP has placed new requirements on the ability of overlay measurement systems to predict real, in-device overlay, taking into account process variations and scanner/lens-induced placement shifts. This is frequently measured by computing the post-modeled measurement residuals after removing all the exposure system correctables. State-of-the-art grating measurements have reduced field-to-field modeled residuals down to 3–4 nm, but DP requires a total overlay budget of 3–4 nm across both exposures.

Additionally, it has been demonstrated that different device feature sizes experience different lateral translations at the wafer plane. To increase precision, previous gratings were larger than the design rule, leading to errors from the difference in lateral shift. Newer grating structures will need to move closer to the design rule size to address this issue.

ISMI is currently producing a test reticle based on alternating phase shift technology, which includes overlay technology test structures focusing on target sizing and overlay accuracy. In addition, ISMI is pursuing next-generation overlay metrology programs, including small in-device targets, diffraction overlay, and secondary-electron-based measurement systems. Work is also proceeding on an overlay RMS to enable high-accuracy in-device measurements.

THIN FILM METROLOGY

As power supply voltages decrease with scaling, enhanced carrier mobility using both compressive and tensile Si/SiGe, as well as SiN stress liners, have become viable options to increase drive current. Material requirements for strained-Si CMOS devices include having uniform SiGe thickness, Ge composition, and strain distribution to maintain consistent across-wafer device performance, as well as low defect density for high minority carrier lifetimes and transconductance and also low surface roughness to minimize the impact of interface scattering on carrier mobilities. Parameters of interest in strained-Si CMOS technology include SiGe and Si channel thickness, Ge composition, strain, dislocation density, interface quality, and roughness. Nondestructive in-line metrology techniques include spectroscopic ellipsometry for film thickness and Ge composition; X-ray reflectivity for thickness, density, and roughness measurements; X-ray fluorescence for Ge composition; UV-Raman spectroscopy for channel strain characterization; in-line TEM for defect detection; and X-ray diffraction for both Ge content and strain measurement. While most of these techniques are well established in the industry, some will require development for application to volume manufacturing. Meanwhile, enhanced mobility devices rely on local strain. The industry needs non-destructive strain metrology with small spot size to achieve the high spatial resolution needed to measure strain on patterned structures.

Ex. 1016, Page 183

Meanwhile, as dimensions shrink and material interaction becomes more subtle, gate spacer profile metrology becomes more challenging. There is a need for high throughput metrology capable of capturing variation in spacer dimension, profile, and material characteristics (Fig. 4). Such metrology must be performed in small areas.

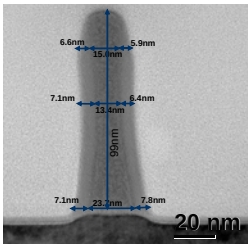


Figure 4: 15 nm gate with 7 nm oxide spacer.

Aggressively scaled gates require ultra-shallow junctions for improved short channel electrostatic control. Metrology capable of measuring dopant activation is critical for process control, as illustrated in Fig. 5. Conventional four-point probe (4PP) methods can damage substrates or produce erroneous measurements due to probe penetration. There is a need for smaller probes that do minimal surface damage or a non-contact technique sensitive to lateral diffusion and depth.

Also, the industry needs metrologies for measuring physical and electrical properties of ultra-thin films and sub-nm interfaces, as shown in Fig. 6. This is essential for producing high-k gate dielectrics with interfaces, which are essential for proper work function tuning through band-gap engineering.

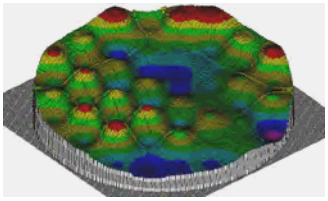


Figure 5: Photoreflectance map of activation of 200eV B implant, 6.5e14 at/cm2, 950ºC anneal (10 sec).

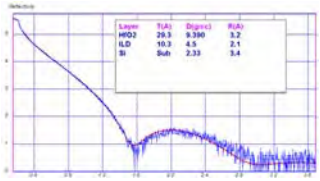


Figure 6: XRR reflectivity spectrum of 30Å HfO2 film.

DEFECT METROLOGY

Unpatterned Defect Detection limits are traditionally measured using polystyrene latex spheres (PSLs). The ITRS standard for particle detection is half of the half-pitch: i.e., 22 nm particles at the 45 nm node, 16 nm particles at 32 nm. Historically, ISMI monitored unpatterned detectability using a combination of test artifacts formed by depositing PSLs on various common film surfaces and by sponsoring theoretical limit studies with NIST and other partners. Theoretical studies funded by ISMI indicate that darkfield limitation of PSL sensitivity on bare Si is 29.8 nm with a signal-to-noise ratio

(SNR) of 10. More aggressive detection data with a lower SNR can push sensitivity down to the mid-20 nm range. These are under optimal conditions; surface haze and roughness have a multiplicative negative effect on detection sensitivity, even on prime silicon. However, at the same time, haze (scattering) measurement from the surface can also be used as a proxy metrology for process monitoring. ISMI is currently investigating the creation of standard haze measurement samples for calibrating the intrinsic roughness measurements used in these systems.

These theoretical studies and practical evaluation using PSLs both indicate that existing unpatterned inspection technology is barely adequate for 45 nm and inadequate for anything smaller. Product development plans from tool makers forecast that shorter wavelength systems will eventually be introduced, which will substantially increase sensitivity to small targets, but this is likely still a few years away.

Meanwhile, PSLs are proving increasingly inadequate as a metric for real unpatterned detection performance. Residues and near-planar defects, which have profiles that look nothing like PSLs, have a significant impact on yield and yet remain difficult to detect. Real-world particles, such as Si and W, also generally have higher refractive indices, resulting in larger sizing. Real particles are also frequently decorated with or embedded into the film material, introducing complications which PSLs cannot easily duplicate. The advantage of PSLs is their excellent traceability, but this may no longer be enough to overcome their increasingly obvious limitations. ISMI is currently investigating the traceability of various process particle species in the hopes of finding a more realistic and powerful set of targets for testing unpatterned inspection.

Patterned Defect Detection ISMI seeks to test the limits of patterned defect inspection by sponsoring theoretical studies in partnership with NIST and with various universities and by producing rigorous practical test artifacts that we use to directly evaluate the performance of available tools. These intentional defect arrays (IDAs) contain SRAM and logic layouts incorporating intentional planar defects ranging in size from 400% to 25% of the half-pitch. They incorporate 14 distinct categories of defects, including shorts, feature extensions, and feature extrusions, allowing ISMI to test sensitivity to many different defect shape and placement configurations. Currently, IDAs for the 45, 32, and 22 nm nodes are each in various stages of production.

Theoretical studies of standard patterned defect detection techniques funded by ISMI indicate that the minimum detection size in brightfield is 20–21 nm, fundamentally limited to 20 nm by wavelength. In darkfield, the minimum detection size is 30 nm, fundamentally limited to 25 nm. In addition, as design rules shrink the wavelengths of light used in these systems treat the device layout as a diffraction grating rather than as a set of distinct features. Reflectance lobing from structures that were previously resolved as straight edges effectively create a new class of nuisance defects that impede the already stretched resolution and noise discrimination abilities of these detection systems.

Although optical tools have historically been the systems most widely used for defect detection, e-beam based systems are enjoying increased popularity due to their ability to surpass the resolution limitations outlined above. However, they suffer from their own problems. First and foremost, they are much slower, a problem that will only get worse as wafer size and active wafer area increases. Furthermore, e-beam inspection requires

Ex. 1016, Page 185

vacuum environments where the reliability of the chamber is considerably less than with optical systems.

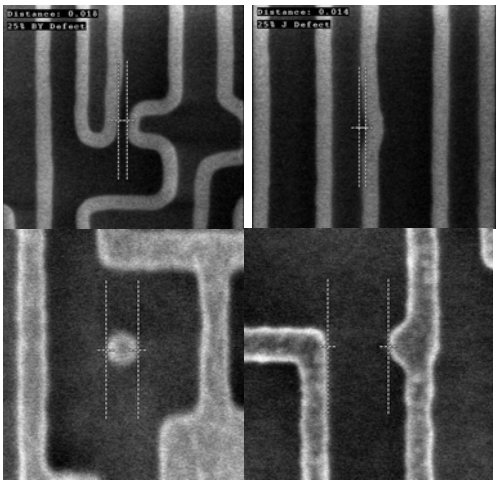


Figure 7: Examples of intentional defects to test defect detection and review tools: (Top-Left) 18-nm bridging defect structure on a 35-nm gate-level structure, (Top-Right) 16 nm edge intrusion in the x direction in a 65-nm MI trench, (Bottom-Left) isolated defect in a 35 nm gate level, and (Bottom-Right) missing line segment defect due to a gap on the 35 nm gate level.

Above and beyond resolution limits, we see several other problems close on the horizon for patterned inspection. Detection modeling requires more realistic 3D construction, for surface defects, embedded defects, and defects within the trench. Also, it is not yet clear that current automated defect classification (ADC) schemes can discriminate real particles from patterned defects. Finally, high aspect ratio inspection (HARI) is becoming a key issue for interconnect defect control. As device margins narrow, contact resistance failure due to residue remaining on the bottom of the contact is becoming a distinct problem, separate from whether the contact is physically open or not.

Defect Review One particular area where shrinking design rules are pushing past the limits of existing technology is the field of defect review. The location of small defects (100 nm and smaller) found by defect detection tools is not precise enough to re-acquire and image the defect without a painstaking search in the imaging tool. Of more fundamental concern is the imaging tool itself. Traditionally, a SEM is used to image the defect, and on-board energy-dispersive X-ray spectroscopy (EDX) is used to analyze its composition. But both these techniques have large interaction volumes that make them fundamentally incapable of resolving defects of 20 nm and below. Analysis is still possible in off-line failure analysis labs using TEM and other techniques, but these are destructive and far too slow to handle the throughput required for in-line review.

Currently, ISMI is engaged in efforts to automate STEM/EELS preparation to elevate throughput and allow its use without destroying the entire wafer. However, other approaches need to be explored as well.

Bevel-Edge Inspection The wafer edge can be a source of several kinds of contamination, including delamination and blisters, which can migrate into the patterned area. This has been exacerbated by the thinning of film layers. However, with the advent

Ex. 1016, Page 186

of immersion lithography, edge inspection is now of critical importance, as the immersion fluid can carry particles once safely isolated in the edge exclusion zone straight into the center of the wafer.

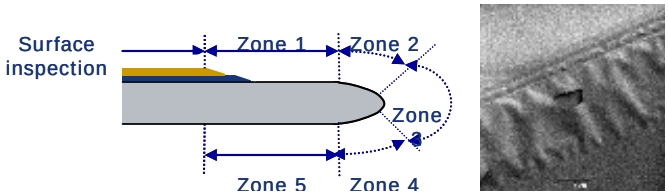


Figure 8: Beveled-edge inspection. (Right) Different zones of the beveled edge. (Left) Blister delamination defect. Image courtesy Infineon.

Currently, most edge inspection tools demonstrate sensitivity down to about 5 m. To detect all edge defects affecting yield, sensitivities down to 1 m will be required. In addition, with immersion, the bottom bevel has become vital, while current tools generally support only the top bevel. Many edge-inspection-capable review SEMs are also unable to image the apex and bottom edge zones. Since the edge area by its nature includes large numbers of nuisance defects, ADC is proving a particularly vital area of advancement. Finally, suppliers must address unacceptably high false alarm ratios and low capture rates when close to the wafer edge exclusion (WEE). ISMI is currently engaged in the manufacture of wafer standards for benchmarking bevel-edge inspection. These will serve as performance targets for qualification, evaluation, and tool monitoring. The lack of accepted standards has to date limited the penetration of wafer inspection into many manufacturers' in-line yield enhancement methodology.

CONCLUSION

The assessments performed by the ISMI metrology program have identified several outstanding issues facing metrology development, in order to achieve full readiness for the 32 nm technology node. As we move forward, ISMI will continue to evaluate the manufacturability of metrology solutions from all these areas and more to help our member companies and the industry at large meet the challenges of these ever-changing and expanding frontiers.

REFERENCES

[1] Bunday, B., Godwin, M., Lipscomb, P., Patel, D., Bishop, M., Allgair, J. & Diebold, A. "Meeting manufacturing metrology challenges at 90 nm and beyond." Micro, August, 2005.

[2] Bunday, B., et al. "Unified Advanced Critical Dimension Scanning Electron Microscope (CD-SEM) Specification for sub-65 nm Technology (2007 Version)." Publicly available at http://www.sematech.org .

[3] The International Technology Roadmap for Semiconductors (San Jose: Semiconductor Industry Association, 2007); available from Internet: http://member.itrs.net.

[4] Orji, G., Dixson, R., Martinez, A., Bunday, B., Allgair, J., and Vorburger, T. "Progress on Implementation of a Reference Measurement System Based on a Critical-Dimension Atomic Force Microscope". J. Micro/Nanolith. MEMS MOEMS, 6(2), 023002, Apr–Jun 2007.

Ex. 1016, Page 187

Advanced Materials Research Center, AMRC, International SEMATECH Manufacturing Initiative, and ISMI are servicemarks of SEMATECH, Inc. SEMATECH and the SEMATECH logo are registered servicemarks of SEMATECH, Inc. All other servicemarks and trademarks are the property of their respective owners.

ECS Transactions, 18 (1) 161-168 (2009)
10.1149/1.3096444 © The Electrochemical Society

Advanced Metrology for Nanoelectronics at the National Institute of Standards and Technology

Joaquin V. Martinez de Pinillos,
Senior Scientist, Office of Microelectronics Programs
Yaw Obeng
Senior Scientist, Office of Microelectronics Programs
Stephen Knight
Director, Office of Microelectronics Programs

A broad range of programs at the National Institute of Standards and Technology address critical metrology and characterization challenges facing the nanoelectronics industry. A brief history of the program will be included. From these programs we will describe some exemplary projects developing new measurement and characterization techniques for nanodevices. These will include, among others, the use of helium ion microscope, spin-transfer microwave nano-oscillators, characterization of graphene, advanced atomic force microscope, and development of a microcalorimeter-based high resolution X-ray analyzer.

Introduction

On May 3, 1900, The US House Committee on Coinage, Weights and Measures met to consider the possibility of creating a national standardization laboratory. Two hours later, the Committee issued a statement to their colleagues in the house that partly said:

> It is therefore the unanimous opinion of your committee that no more essential aid could be given to manufacturing, commerce, the makers of scientific apparatus, the scientific work of the Government, of schools, colleges, and universities than by the establishment of the institution proposed by this bill (1).

In March 3, 1901, the bill founding the National Bureau of Standards (NBS) passed both houses of Congress (2).

In 1988, the U.S. Congress changed the name from NBS to the National Institute of Standards and Technology (NIST) with four parts to accomplish its functions. They are the following:

- the **NIST Laboratories**, conducting research that advances the nation's technology infrastructure and is needed by U.S. industry to continually improve products and services;
- the **Baldrige National Quality Program**, which promotes performance excellence among U.S. manufacturers, service companies, educational institutions, and health care providers; conducts outreach programs and manages the annual Malcolm Baldrige National Quality Award which recognizes performance excellence and quality achievement;

Ex. 1016, Page 189

- the **Hollings Manufacturing Extension Partnership**, a nationwide network of local centers offering technical and business assistance to smaller manufacturers; and
- the **Advanced Technology Program**, which accelerates the development of innovative technologies for broad national benefit by co-funding R&D partnerships with the private sector.

NIST's mission is to develop and promote measurement, standards, and technology to enhance productivity, facilitate trade, and improve the quality of life.

Presently, NIST consists of 8 measurement laboratories, the Manufacturing Extension Program (MEP), and the Baldrige National Quality Program. There are approximately 2,800 employees in Gaithersburg, Maryland, Boulder, Colorado, and Charleston, South Carolina, 1,800 guest researchers from all over the world, and 850 users of facilities. The 9 measurement laboratories are: Building and Fire Research, Chemical Science and Technology, Electronics and Electrical Engineering, Manufacturing Engineering, Materials Science and Engineering, Physics, Technology Services, Information Technology, and the Center for NanoScience and Technology There are three Nobel laureates in Physics working at NIST: Bill Phillips (1997), Eric Cornell (2001), and Jan Hall (2005).

NIST serves a broad base of customers in environmental technologies, transportation, pharmaceuticals, law enforcement, food and nutrition, biotechnology, computer software and equipment, construction, manufacturing, and microelectronics. In serving these customers, NIST partners with companies, national laboratories, academia, and agencies of the state and federal governments.

The Microelectronics Industry

In 1990, the U.S. Congress created the National Semiconductor Metrology Program (NSMP) to address the metrology needs of the semiconductor industry. NIST, in turn, created the Office of Microelectronics Programs (OMP) to manage these programs internally. OMP funds researchers in different NIST laboratories to address semiconductor industry needs. Personnel from OMP spend significant time working with semiconductor companies, e.g., IBM, Intel and Texas Instruments (TI); consortia, e.g., the Semiconductor Research Corporation (SRC), the Semiconductor Equipment and Materials International (SEMI) and SEMATECH; and academia. These interactions allow the NIST researchers to stay current on the fast changing, increasingly demanding, industry needs. In 2007, NIST joined the Nanotechnology Research Initiative (NRI), part of the Semiconductor Research Corporation (SRC), by contributing $2.8M a year for five years starting in 2007.

OMP has research programs in the following areas: Lithography, Critical Dimension and Overlay, Front-End Processing, Interconnect and Packaging, Process, Analysis Tools and Techniques, Device Design and Characterization, System Design and Test, and Manufacturing Support. There are several projects within each program area being conducted throughout NIST.

Ex. 1016, Page 190

The Semiconductor Industry Association (SIA) has developed a way of communicating with research institutions to inform them of their needs in the future called the International Technology Roadmap for Semiconductors (ITRS) which is published every two years with an update on the intermediate years. Table I shows one of the tables in the 2007 ITRS[3] indicating the critical dimensions (CD) expected during the forthcoming years for the different products that the industry manufactures. CD is the minimum width of the lines printed on a chip. There are different definitions of CD depending on the type of chip. Table II[3] shows a typical depiction of some properties that, according to the ITRS, will be of important in future years for different technologies. The ITRS uses a color code to identify the state of the technology that addresses those issues: white indicates that manufacturable solutions exist and are being optimized; yellow, manufacturable solutions are known; and red, manufacturable solutions are not known. When interim solutions are known, a white background with yellow lines and a red rhombus appears.

We will now address in more detail some of the OMP supported projects.

Discussion of Projects

Spintronics

As device dimensions shrink to below about 200 nm, the spin degree of freedom allows for new ways to control and manipulate magnetism. What could previously be done only with applied magnetic fields can now be done with spin polarized currents flowing through device structures. This has allowed the investigation and development of a number of new device concepts. The fundamental reason for this is that in nanostructured devices the angular momentum associated with the electron spins in a current can now be coherently transferred to the magnetization state of a ferromagnetic material.

NIST is developing the metrologies to understand these interactions with a particular focus on how these new effects can be harnessed for new devices. We are focusing on how these spin interactions can be used to generate new kinds of nanoscale microwave devices and to control magnetization for information storage in magnetic RAM, as well as more “far out” ideas to develop spin batteries, spin-voltage readout schemes, and spin-based logic concepts. See Fig.1.

Other aspects that NIST is looking at are architectures that use spinwaves to transmit and receive information. One example of this is shown in Fig. 2 where we fabricated two spin-based high frequency oscillators about 500 nm apart on a magnetic film.

The most mature field where these new spin interactions have found potential application is in the area of magnetic random access memory (MRAM). The basic idea of MRAM is shown in the upper left of Fig. 3. The simplest incarnation of MRAM has two magnetic layers. One of these is engineered to be fixed (as the reference layer) and the other is made to point left or right and acts as the storage layer.

Ex. 1016, Page 191

Carbon Nanotubes and Graphene

Carbon based materials have interested many scientists and engineers as potential materials for future semiconductor manufacturing.

Table 1a&bProduct Generations and Chip Size Model Technology Trend Targets

	Year of Production	*2007*	*2008*	*2009*	*2010*	*2011*
	Flash ½ Pitch (nm) (un-contacted Poly)(f)	**54**	**45**	**40**	**36**	**32**
WAS	*DRAM ½ Pitch (nm) (contacted)*	*65*	*57*	*50*	*45*	*40*
IS	***DRAM ½ Pitch (nm) (contacted)***	***68***	***59***	***52***	*45*	*40*
IS	*MPU/ASIC Metal 1 (M1) ½ Pitch (nm)*	*68*	*59*	*52*	*45*	*40*
WAS	***MPU Printed Gate Length (nm)***	**42**	**38**	**34**	**30**	**27**
IS	***MPU Printed Gate Length (GLpr) (nm) ††***	**54**	**47**	**41**	**35**	**31**
WAS	***MPU Physical Gate Length (GLph) (nm)***	**25**	**23**	**20**	**18**	**16**
IS	***MPU Physical Gate Length (GLph) (nm)***	**32**	**29**	**27**	**24**	**22**
WAS	***ASIC/Low Operating Power Printed Gate Length (nm) ††***	**54**	**48**	**42**	**38**	**34**
IS	***ASIC/Low Operating Power Printed Gate Length (nm) ††***	**64**	**54**	**47**	**41**	**35**
WAS	***ASIC/Low Operating Power Physical Gate Length (nm)***	**32**	**28**	**25**	**23**	**20**
IS	***ASIC/Low Operating Power Physical Gate Length (nm)***	**38**	**32**	**29**	**27**	**24**
ADD	***ASIC/Low Standby Power Physical Gate Length (nm)***	**45**	**38**	**32**	**29**	**27**
ADD	***MPU Etch Ratio GLpr/GLph (nm)***	**1.6818**	**1.6039**	**1.5296**	**1.4588**	**1.4237**

Table I. Part of the Product Generation and Chip-Size Model Technology Trends-Near Term Years from the ITRS 2008 update to the 2007 version.

Table MET2 *Metrology Technology Requirements*

Year of Production	*2007*	*2008*	*2009*	*2010*	*2011*
Microscopy					
Inline, nondestructive microscopy process resolution (nm) for P/T=0.1	**0.22**	**0.2**	**0.18**	**0.16**	**0.14**
Microscopy capable of measurement of patterned wafers having maximum aspect ratio/diameter (nm) (DRAM contacts) [A]	**16** **76**	**17** **67**	**17** **60**	**>20** **50**	**>20** **40**
Materials and Contamination Characterization					
Real particle detection limit (nm) [B]	**25**	**22**	**20**	**18**	**16**
Minimum particle size for compositional analysis (dense lines on patterned wafers) (nm)	**22**	**19**	**17**	**15**	**13**
Specification limit of total surface contamination for critical GOI surface materials (atoms/cm^2) [C]	**5.00E+09**	**5.00E+09**	**5.00E+09**	**5.00E+09**	**5.00E+09**
Surface detection limits for individual elements for critical GOI elements (atoms/cm^2) with signal-to-noise ratio of 3:1 for each element	**5.00E+08**	**5.00E+08**	**5.00E+08**	**5.00E+08**	**5.00E+08**

Manufacturable solutions exist, and are being optimized

Manufacturable solutions are known

Interim solutions are known

Manufacturable solutions are NOT known

Table II. Part of the Metrology Requirements Table from the 2007 ITRS.

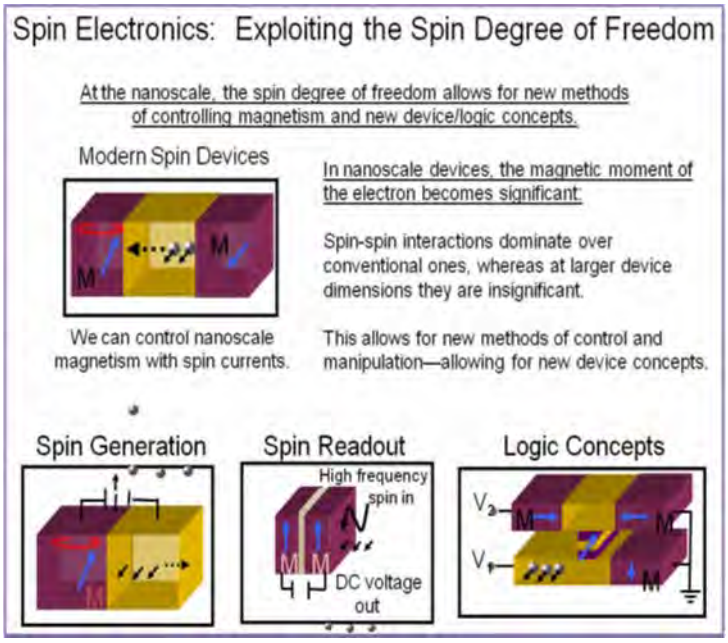


Figure 1. (top) This shows the basic concept behind the "spin transfer" effect: that information from one magnetic layer can be coherently transferred to another through spin polarized electroncs. This provides a new means of controlling and manipulating magnetism. (bottom) This effect holds the promise to allow the development of spin batteries, efficient readout of high frequency information, and new spin-based logic. The basic concepts behind these devices have been experimentally demonstrated. The focus of the NIST spin electronics program is to investigate their performance in practical device architectures.

Graphene can be thought of as a 2-dimensional structure composed of rings of six carbon atoms each. They are related to graphite, but graphite is a 3-dimensional structure that can be thought of as layers of graphene superimposed on one another. So, technically, graphene is a monolayer material, but bi-layers, tri-layers, etc. have been shown to have electrical properties similar to monolayer graphene. NIST is developing metrology techniques to measure accurately the properties of graphene.

Carbon nanotubes (CNTs) can be thought of as rolled sheets of graphene. CNTs can be either single-walled (SWNT) or multi-walled (MWNT). They have properties of great interest, and measuring those properties has been a challenge because CNTs are usually prepared under which impurities and non-CNTs carbon compounds contaminate the CNTs. Some NIST researchers have developed a laser cleaning system that is discussed in Figs. 4 through 7.

Other applications will be discussed during the presentation.

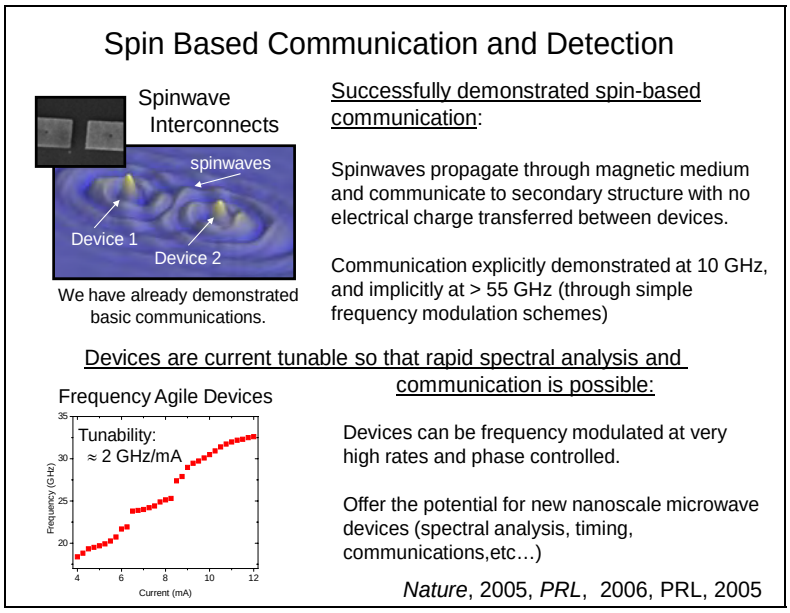


Figure 2. (top) Micomagnetic simulation of spin-based communication between two spin oscillators. The devices transfer information without any electrical charge transfer between them. This experimental demonstration was performed at 10 GHz. (bottom) Output from a single device showing that the oscillator frequency can be tuned between 20 GHz and 30 GHz by varying the current through the device. More recent results show devices working as high as 55 GHz.

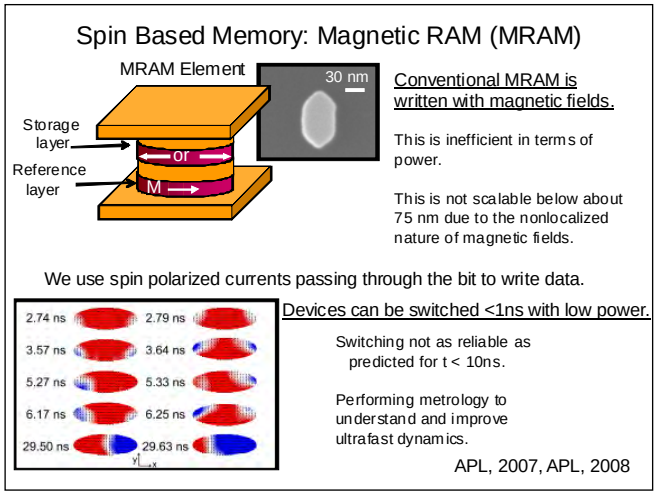


Figure 3. (top) Schematic of an MRAM device along with an SEM view of an indivual MRAM element (patterned at Hitachi Global Storage Technologies (HGST)). (bottom) Micromagnetic simulations (from a simulator developed at NIST (OOMMF)) suggest that the details of the reversal (switching) process are more complicated than typically assumed. Understanding the details of the switching process is expected to be crucial to the future development of MRAM.

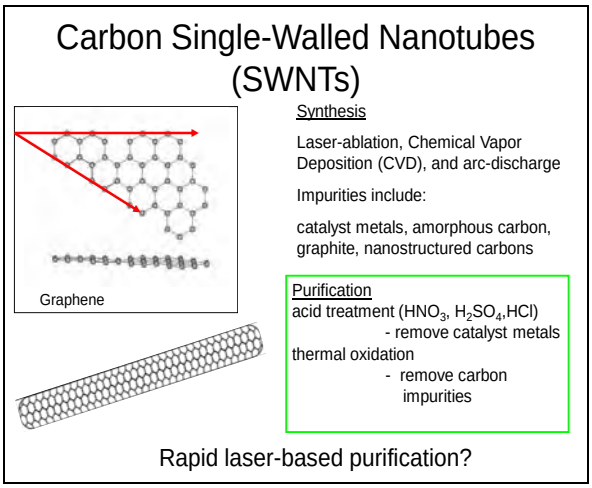


Figure 4. Description of the preparation and purification on single-walled carbon nanotubes.

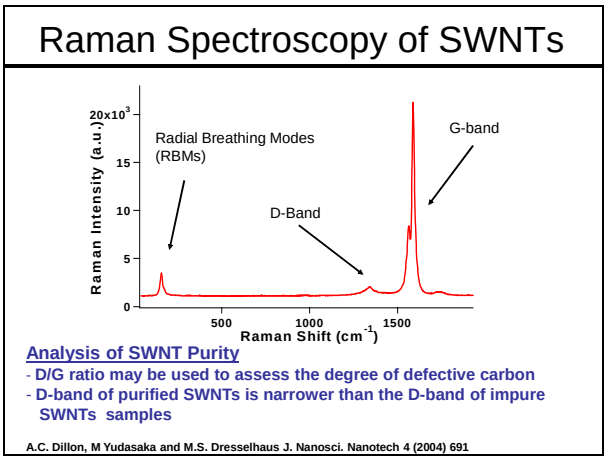


Figure 5. Raman spectrum of SWCNT

Conclusions

NIST's support has been and still is critical for the metrology needs of the semiconductor industry. Programs at NIST that are directed to this industry have provided needed techniques and methodologies that have allowed the industry to progress rapidly into areas of smaller dimensions and incorporation of other technologies in the same device; e.g., telephones with connection to the internet and cameras.

Acknowledgements

We would like to acknowledge Dr. William Rippard and Dr. John Lehman from NIST for help with the figures and technical discussions.

Ex. 1016, Page 195

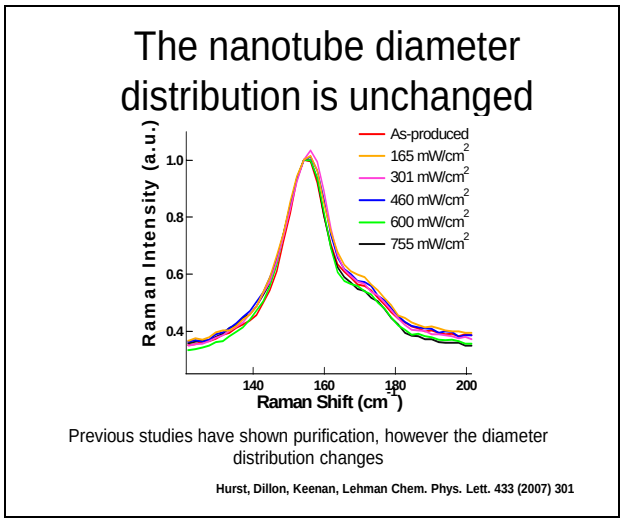


Figure 6. The CNT diameter is unaffected by the laser treatments.

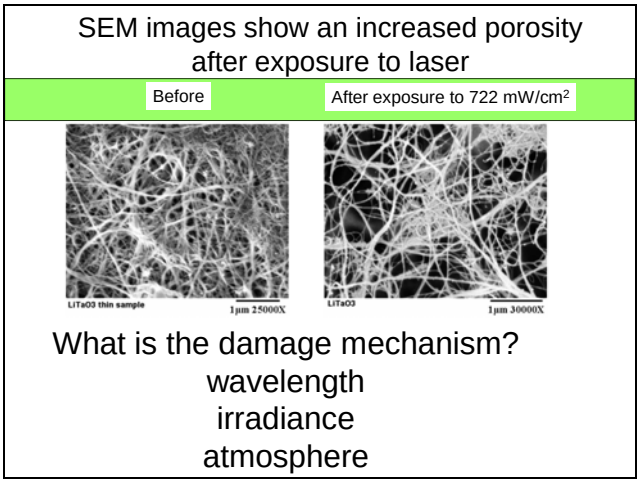


Figure 7. Laser exposure enhances the porosity of the matrix where the CNTs are embedded.

References

1. H.R. 1452, "National Standardizing Bureau," Cong., 1st sess., May 14, 1900 (US House Reports serial 4026, vol. 6, 1899 – 1900). This is the inscription over the new Bureau laboratories in Gaithersburg, MD.
2. Cochrane, R. C., "Measures for Progress: A History of the National Bureau of Standards," U.S. Department of Commerce, 1966.
3. International Technology Roadmap for Semiconductors, Semiconductor Industry Association, 2007.

Ex. 1016, Page 196

ECS Transactions, 18 (1) 169-175 (2009)
10.1149/1.3096445 © The Electrochemical Society

Advanced 3D-AFM Metrology for Sidewall Spacers

T. Bao, V. Ukraintsev, and D. Dawson

Veeco Instruments Inc, 112 Robin Hill Rd, Santa Barbara, CA 93117, USA

Gate spacer engineering has become one of the primary concerns in dimensional metrology for semiconductor manufacturing process control. In composite spacers, deposition and etch of each spacer determine where the transistor source/drain implants occur. The sidewall spacer thickness and profile, especially at the feature bottom, is critical to be characterized and controlled by applicable 3D metrology. This paper discusses recent advances in 3D atomic force microscopy (3D-AFM) that solve the specialized characterization needs for critical sidewall spacer geometry controls, including multiple spacer thickness and nitride spacer pulldown. 3D-AFM metrology measures with greater accuracy such typical gate spacer parameters as linewidth, height, pitch, sidewall profile, sidewall angle (SWA), line edge roughness (LER), and line width variation (LWV), and sidewall roughness (SWR).

Introduction

Sidewall spacers are used for precise MOSFET engineering, and they directly define channel length, junction geometry and abruptness, therefore the final transistor electrical characteristics. Spacer engineering has advanced to a level where conventional CD-SEM and optical scatterometry (OCD) face fundamental limitations. Since advanced logic and memory semiconductor device manufacturing strictly adheres to the ITRS roadmap (see Table 1), manufacturers have turned to advanced 3D metrology to control the feature shape dimensions for 65 and 45nm node patterns created by microlithography processes.

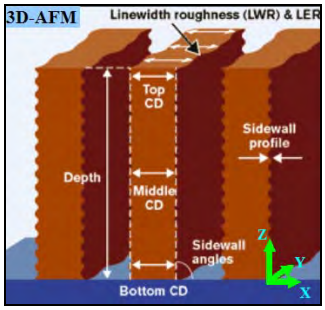
Table 1. ITRS semiconductor roadmap 2007.

Technology Node (nm)						
Year in Production	2007	2008	2009	2010	2011	2012
ASIC M1-1/2 Pitch	68	59	52	45	40	36
ASIC Gate Length	25	22	20	18	16	14
DRAM M1-1/2 Pitch	65	57	50	45	40	36
FLASH M1-1/2 Pitch	54	45	40	36	32	28

Table 2. Gate metrology applications of AFM.

Gate Module	Measurement	Metrology Usage
Films (diele./poly)	film roughness	engineering characterization
Gate Resist	resist height, CD, SWA, LWR, LER, SWR	in-line process control and reference metrology, proces development and tool qualification
Gate Etch	gate height, CD, SWA, LWR, LER, SWR	
Gate Spacers	each spacer height, SW thickness, SWA	
Spacer Pulldown	pulldown height	
Dual Stressor Liner	liner bottom CD, nitride residual	
Recessed Strain SiGe	recess depth, lateral undercut, channel CD	engineering characterization
Fin-FET	fin/gate height, CD	engineering characterization

Figure 1. Critical characteristics of 3D gate control.



Multiple spacers, deposition and etch of each spacer determines where S/D implants occurs. The sidewall spacer thickness and profile especially at feature bottom is critical to be characterized and controlled by applicable 3D metrology.

Ex. 1016, Page 197

3D-AFM metrology measures with greater accuracy such typical gate parameters as linewidth, height, pitch, sidewall profile, sidewall angle (SWA), line edge roughness (LER), and line width variation (LWV), and sidewall roughness (SWR). See Table 2 and Figure 1 for a detailed explanation of measured geometry parameters. In addition, it can be used for complimentary reference and calibration metrology, inline process monitoring, or engineering characterization.

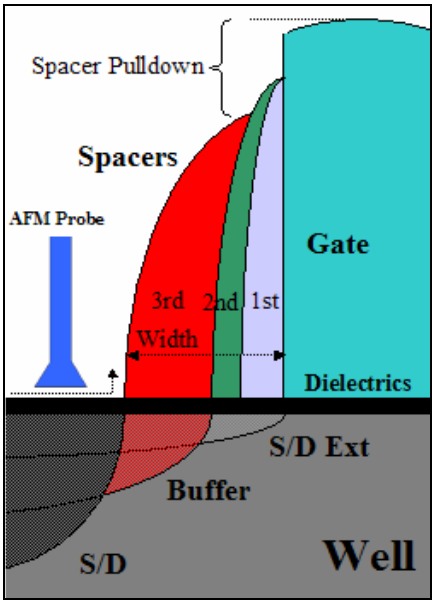
Spacer Profile Control

Shown in Figure 2, each sidewall spacer is a thin dielectric film formed by highly anisotropic reactive ion etch (RIE) on gate sidewalls. The spacers provide lateral offsets for sequential implants during source-drain (S/D) junction engineering in the transistor channel. The process flow alternates between several spacer formation and implant steps. The sidewall spacers are crucial for proper control of lateral positions and overlap of sequential doping implants (shallow S/D extension, halo and deep S/D) with respect to the edge of the transistor gate. The spacers determine the channel length, exact junction geometry and abruptness that are directly related to the final transistor electrical performance.

Along with the shrinking gate and channel length, the depth of the S/D extension (SDE) junction needs to be reduced accordingly to maintain and even further improve transistor on/off performance. However, reduction of SDE depth noticeably increases the transistor series resistance (thus reducing drive current). A variety of spacer structures can be used between multiple implant processes to tailor the S/D junction to maximize drive current while maintaining low parasitic capacitance of the transistor.[1]

As shown in Figure 2, the first offset spacer is formed before (pMOS) or after (nMOS) the lightly doped SDE junctions are formed, but before halo implants are put in the channel. Then second and third nitride spacers are formed for heavier buffer and S/D implants, respectively. The details of transistor integration may vary between different manufacturers but the general idea remains the same.

Figure 2. Schematics of multiple sidewall spacers and S/D junctions (showing one side only. Actual spacer structures may vary).



The lateral dimension of corresponding spacer directly dictates the position of shallow SDE, buffer, and deep S/D junctions. These junctions define critical electric characteristics of the transistor (threshold voltage, drive current, channel resistance/leakage, load capacitance, and short channel effects).

By tuning deposited film thicknesses (CD), RIE processing (CD and profile), and optimizing implant dose/energy, one can achieve the desired SDE, buffer and S/D design. Therefore, the sidewall spacer CD and profile control is critical. The profile and bottom width of each spacer needs to be controlled to ensure consistent transistor electrical performance lot-to-lot, wafer-to-wafer, across wafer and across integrated circuit (IC).

Ex. 1016, Page 198

With proper gate integration and spacer process control, one can significantly improve transistor drive current and load capacitance, and therefore speed. For example, 25% improvement of pMOS saturation drain-current has been achieved with the triple spacer transistor design over the single spacer structure.[1]

Unfortunately, transistor electrical performance is evaluated only when the first interconnect is completed. This happens weeks after implant processing. Providing in-line spacer process control much earlier in the flow provides an effective way to prevent unnecessary yield loss related to gate integration.

Spacer Metrology Challenges

Spacer thickness and sidewall profile are important parameters to control. The spacer thickness and profile measurements are extremely challenging and continue to be problematic for most available thin film metrologies, including x-ray, optics, acoustic and e-beam based technologies. Spacer engineering requires a true 3D characterization solution for sidewall thickness, sidewall profile, and spacer pulldown height. Table 3 compares various dimensional metrology techniques available today.

Table 3. Advantages of 3D-AFM in comparison with other metrology techniques.

	CD-SEM	OCD	Dual-Beam	TEM	Next-Gen 3D-AFM
Area of use	Process control OPC & RET modeling Engineering	Process control Litho evaluation	Failure analysis Engineering	Reference tool Failure analysis Engineering	Accurate reference tool Engineering Process control OPC & RET modeling
Sample interaction	Shrinking Charging Carbonization	Bleaching (especialy for spectral tools)	Destructive Galium contamination Shrinking	Destructive Shrinking Warping	Non-destructive No contamination No sample bias
3D Capability	2D (XY)	2D (XZ)	2D (XZ)	2D (XZ)	3D (XYZ)
Cost per site	low	lowest	high	very high	moderate
MAM time	4-5 sec	2-4 sec	30-90 sec	2-4 hours	10-40 sec
Precision (gate CD)	0.7 nm (3s)	0.4 nm (3s)	ca. 1-2 nm (3s)	ca. 1-2 nm (3s)	1.0 nm (3s)
Uncertainty (gate CD)	ca. 1.1 nm (3s)	ca. 1.0 nm (3s)	ca. 2.4 nm (3s)	ca. 2.2 nm (3s)	ca. 0.5 nm (3s)
Uncertainty (gate SWA)	ca. 0.8 deg (3s)	0.7 deg (3s)	ca. 0.6 deg (3s)	ca. 0.6 deg (3s)	0.2 deg (3s)
Limitations	< 15nm CD Variable bias for 2D & dense features Material & profile bias dependence STI CD variable bias starting 65nm	Gate SWA bias starting 45nm Bottom CD bias starting 45nm STI CD variable bias starting 65nm Model parms cross-correlation	Limited sampling Shrinking Charging Sampling uncertainty	Limited sampling Shrinking Warping Sampling uncertainty	< 50nm spaces 10nm blind bottom zone

The lack of information in the third, vertical dimension prevents reliable measurements of spacer bottom, sidewall profile, and vertical pulldown by CD-SEM. SEM is also a subject of CD bias variation caused by electrostatics, material differences, and proximity effects.

Similarly, OCD is affected by CD bias variation caused by changes of sample geometry, film stacks, and material optical properties. In addition, the multi-layered spacer stack and geometry are too complex for OCD to generate accurate models. OCD can work only on the periodic gratings located in the wafer scribe areas. This also implies that OCD provides indirect sidewall spacer monitoring of actual in-die circuits. Correlation between scribe and in-die data should be constantly monitored in changing process conditions.[2]

There is a need for reference measurements to validate accuracy for both CD-SEM and OCD. Traditional reference methods, such as cross-sectional SEM and TEM, are destructive, expensive and can't be implemented in-line. More importantly, the high local sampling uncertainty associated with a single cross-section measurement makes both XSEM and TEM unacceptable as a reference for accuracy validation beyond a few nanometers. There is simply too much measurement uncertainty in these techniques. It is no accident that ITRS 2007 has proposed a new metrology quality metric, measurement uncertainty (MU), for evaluation of in-line CD metrology. Sampling uncertainty is a component of MU and, therefore, must be minimized.

Ex. 1016, Page 199

3D-AFM for Gate Spacer Metrology

MU, as defined by 2007 ITRS, includes reproducibility or precision, sampling uncertainty, tool-to-tool matching and other inaccuracy components. Sample-to-sample bias variation is one of the major contributors to the inaccuracy component. 3D-AFM is capable of delivering the lowest uncertainty of non-destructive measurements. Thus, it can be used as a reference metrology system to measure absolute dimensions of samples representative to specific technology in determining the sample-to-sample bias uncertainty component of the metrology tool under test. Currently, 3D-AFM is the only practical way to determine sample dependent bias, and to validate and calibrate in-line CD-SEM and OCD.

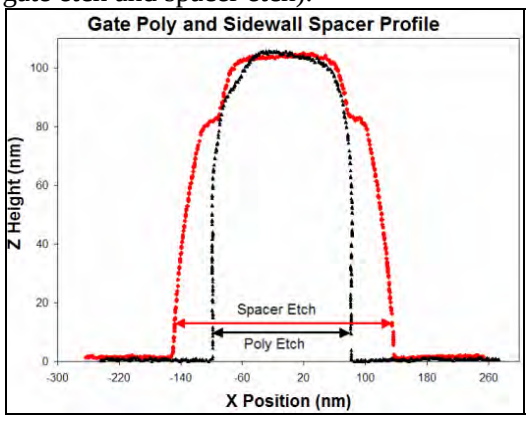
3D-AFM measurements are independent of sample, proximity, and feature density. It gives the absolute accuracy with minimal CD measurement uncertainty. 3D-AFM provides non-destructive measurement with nanometer precision and accuracy, fully automated data analysis, and high productivity of up to 10-40 sec/site. This makes 3D-AFM a good metrology tool for in-line monitoring and engineering analysis of sidewall spacers without developing OCD models or scraping wafers for X-SEM or TEM analysis.

Since 3D-AFM does not need any special structure, it is also a true in-die metrology that measures discrete and isolated gates such as ring oscillator transistors. Therefore, it could replace OCD and SEM as a preferred in-line metrology. In any case, it is a complementary in-line metrology to OCD and CD-SEM and a replacement reference tool for X-SEM and TEM.

Measuring Sidewall Thickness

Due to the unique capability of pattern recognition, 3D-AFM can precisely place the tip at the same spot on the same wafer through sequential process steps. Metrologists use AFM to scan the same gate line after gate etch, and later after gate spacer etch, to obtain CD and profile data for each step (see Figure 3). The difference conveniently gives the spacer sidewall thickness and profile. The nature of direct measurement with absolute accuracy enables the user to take the guesswork out of optical metrology and also overcomes the CD-SEM weakness in sidewall bottom CD.

Figure 3. Using AFM for sidewall spacer thickness (delta bottom linewidth between gate etch and spacer etch).



Validating OCD Models

The technology at 65nm node and below imposes very stringent (comparable to the gate CD) tolerances for spacer thickness variation. Because of its obvious advantages in speed, cost and precision, OCD is a good candidate. Once again, accuracy and TMU of OCD data need to be verified. Traditionally, TEM has been widely used for reference and calibration. Now however, 3D-AFM is a better replacement for TEM, with better MU of spacer thickness measurements to verify and calibrate sidewall spacer OCD models. In addition, it can quickly and non-destructively produce a large amount of valuable data.

Ex. 1016, Page 200

Figure 4 shows results of OCD model verification by using 3D-AFM as a reference tool. Here, the apparent discrepancy between OCD and AFM indicates further OCD model development is required to improve accuracy.

Figure 4. 3D-AFM data confirm that the OCD model is not robust and needs improvement.[3]

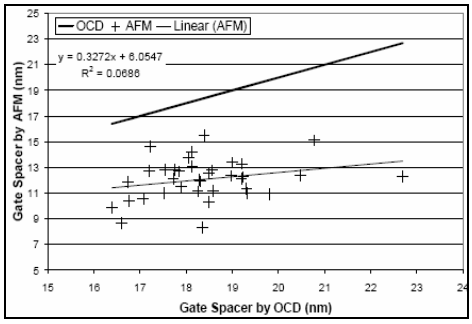
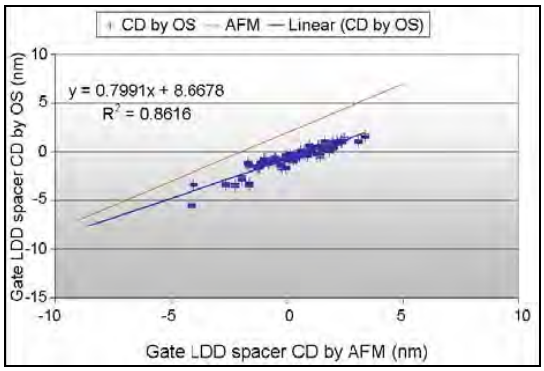
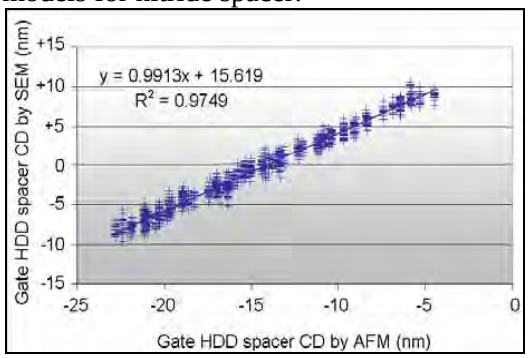


Figure 5. 3D-AFM validates the OCD models for offset spacer.[4]



To verify accuracy of the 1st offset gate spacer OCD metrology, the total gate bottom CD was calibrated by 3D-AFM at the post spacer etching process level. Figure 5 presents the correlation between the OCD and reference data. The data shows a correlation of 0.86 and a slope of 0.80. The fleet bias for the bottom gate CD at this level was -2.8nm. The fleet TMU was estimated as 0.70nm (3σ) using uncertainty of the reference bottom gate CD of 0.95nm (3σ). The OCD single tool and fleet precisions for the bottom gate CD. were 0.32nm (3σ) and 0.41nm (3σ), respectively. Corresponding OCD single tool and fleet precisions for the thickness of offset spacer were 0.17nm (3σ) and 0.21nm (3σ), respectively. The OCD provides a satisfactory fleet TMU (TMU/T is below 0.2) and acceptable accuracy for the bottom gate CD control at offset spacer etchingThe nitride spacer (2nd) stack and geometry is too complex for OS modeling. Therefore, SEM is used to control the total gate bottom CD and spacer thickness at the nitride spacer etching process level. Two exposure matrix (EM) wafers with 70 pre-calibrated sites (260nm pitch OCD structure) were used for measurement system analysis. The wafers were measured more than seven days on three baseline SEM tools. Figure 6 presents the correlation between the SEM and reference data. The data shows a correlation of 0.99 and a slope of 0.99. The fleet bias for the bottom gate CD at the nitride spacer etching was 14.2nm (cf. to SEM bias of 8.8nm at the gate etching). The fleet TMU was estimated as 2.1nm (3σ) using uncertainty of reference data of 1.1nm (3σ). The OCD single tool and fleet precisions for the bottom gate CD were 1.3nm (3σ) and 2.0nm (3σ), respectively. The SEM fleet TMU/T ratio for the nitride sidewall spacer control was at an acceptable level below 0.2. Precision of the 3D-AFM measurements was determined for every analyzed process level based on repeated measurements conducted over several days. The NIST traceable standards were used to verify 3D-AFM line CD and XYZ scale calibrations. The NIST traceability was established for CD metrology at every critical process level for the entire technology. As can be clearly seen here, 3D-AFM fully succeeds in validating the OCD models.

Figure 6. 3D-AFM validates the OCD models for nitride spacer.[4]



Evaluating Spacer Profiles

Process and integration engineers also want to know the evolution of differential spacers through sequential steps of the gate processing. To get such information, one would like to measure exactly the same gate several times through the processing to evaluate the effects of etch and cleaning processes on the gate profiles. Ideally, a simple and direct measurement would be possible that would not require simulations or scrapping of wafers, as is the case with X-SEM or TEM.

3D-AFM is ideally suited for this type of work. It is important to note that sample-to-sample bias variation is minimal for 3D-AFM and, therefore, no bias correction is needed through the measurement. Figure 7 presents an example where one gate with about an 80nm-thick spacer went through the series of process steps, including 2 RIE and 1 post-RIE cleanup. Only the first RIE has modified the gate CD and profile significantly. The cleanup and the second RIE had only a minor effect on gate profile and CD. Figure 8 shows a direct correlation of offset thickness with the transistor current. Source drain implants can be adjusted as a function of inline offset spacer thickness measurements at post offset etch.

Figure 7. Sidewall profiles of differential spacers throughout the process flow.

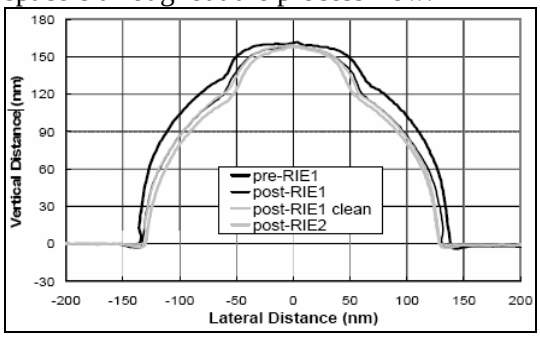
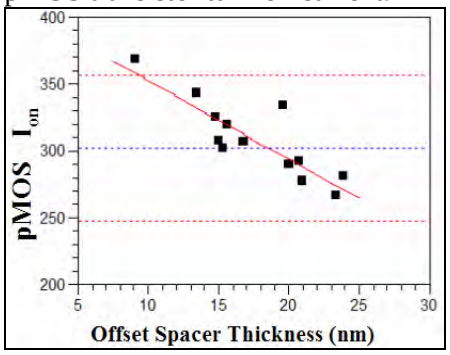


Figure 8. Offset spacer thickness defines the pMOS transistor turn-on current.



Measuring Nitride Spacer Pulldown

In the final spacer formation, anisotropic RIE etch preferentially removes nitride films from the flat surface faster than the sidewall surface, leaving behind the spacer film on the gate sidewalls. The salicidation process will form nickel silicide films over all the exposed silicon in gate electrode and source/drain area to form transistor contacts.

The sidewall spacer provides a uniform coverage over gate sidewall against salicidation. However, if RIE etching is excessive, the spacer can suffer top loss, causing the top of the spacers to be pulled down lower than the gate electrode top (see Figure 2). As a result, the gate top corner sidewalls become exposed. The uncovered gate material can form nickel silicide during the subsequent salicidation process. This reduces the distance between source/drain silicide and gate sidewall silicide, and there is an increased chance of gate leakage. In addition, formation of silicide stringers and ribbons results in device shorts. Therefore, there is much interest in accurately measuring and controlling nitride spacer pulldown height (NSPD).

Unlike CD-SEM (no vertical data) and OCD (model inaccuracy), 3D-AFM provides a direct, reliable, and nondestructive measurement for NSPD height. The field 3D-AFM data showed a repeatability of 2nm (3σ) can be achieved for NSPD. Figure 9 shows X-

Ex. 1016, Page 202

SEM and 3D-AFM agree well on the typical NSPD measurement. However, X-SEM only measures at one cleavage. For 3D-AFM, the shown profile in Figure 9 is the average of 16 scan lines (or cleavages) along the gate length. The data is taken in less than 40 seconds. This further exemplifies the benefit of 3D-AFM in providing large sampling to reduce measurement uncertainty.

Figure 9. Nitride spacer pulldown height measured by AFM agrees with X-SEM.

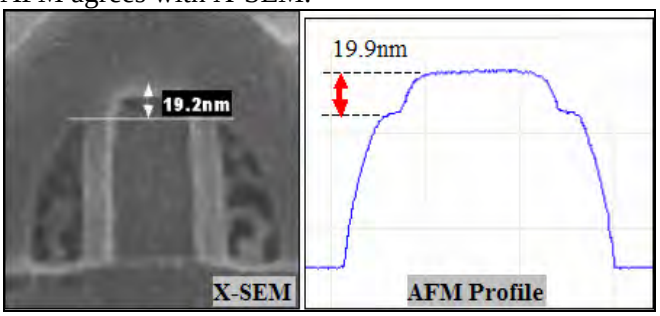


Figure 10. Validation of AFM NSPD measurements by X-SEM.

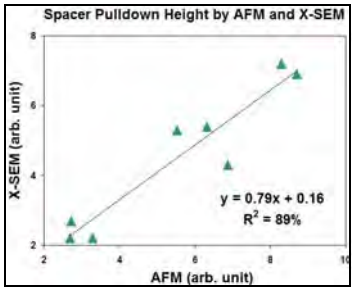
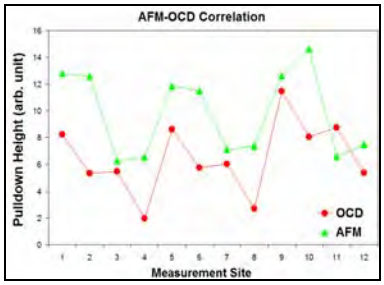


Figure 11. Correlation of AFM and OCD for NSPD data. \



To further confirm 3D-AFM data, Figure 10 shows a good correlation of AFM with X-SEM measurements across multiple sites for NSPD. The data highlights AFM can be used to replace destructive X-SEM as a reference tool for OCD to validate model simulations. In a separate effort, the OCD data in Figure 11 shows an apparent discrepancy with 3D-AFM, and the OCD models are inadequate.

Conclusion

New-generation 3D-AFM has a demonstrated capability for stand-alone production monitoring as well as a complementary solution for challenges facing the modern CD-SEM and OCD. 3D-AFM is a crucial reference, inline monitoring, or engineering characterization tool for advanced spacer engineering. Especially, as a production worthy reference tool, 3D-AFM provides bias free measurements to evaluate the TMU, verify and calibrate inline CD-SEM or OCD metrology systems, for sidewall spacer thickness and spacer pulldown height.

References

1. M. Horstmann, D. Greenlaw, et al., “Advanced Transistor Structures for High Performance Microprocessors,” IEEE International Conference on Integrated Circuit Design and Technology, p. 65 (2004).
2. J. Allgair, B. Bunday, “A Review of Scatterometry for Three-Dimensional Semiconductor Feature Analysis,” Future Fab Intl. V.19 (June 2005).
3. V. Ukraintsev, et al., “The role of AFM in Semiconductor Technology Development: The 65nm Technology Node and Beyond,” Proceedings of SPIE V. 5752, p. 127–39 (May 2005).
4. V. Ukraintsev, et al., “Transition from Precise to Accurate Critical Dimension Metrology,” Proceedings of SPIE V. 6518 (April 2007).

Ex. 1016, Page 203

ECS Transactions, 18 (1) 177-182 (2009)
10.1149/1.3096446 © The Electrochemical Society

High Yield Patterning with Simplified DFM Rules in Sub-100nm Process Technologies

Li-Fu Chang, Chang IL Choi, Min-Hwa Chi

Semiconductor Manufacturing International Corporation, 18 Zhangjiang Road, Pudong New Area, Shanghai 201203, People's Republic of China

Traditional DFM methods of sub-100nm technologies have emphasized on offering of optical proximity effect driven DFM rules for designers to reduce shape variations of the conducting patterns. This offering has been complicating over the rolling-out of new technologies in the past. The complication is most severe with the semiconductor foundries, which manufacture many different designs with the same set of tools. However, not only are the DFM rules an increasing work load, but recent reports also found new manufacturing failures are produced along with the change of design required by the DFM rules. A new DFM methodology that eliminates such problems is therefore necessary. In this paper, we show how we have achieved high yield in patterning with a significantly simplified set of DFM rules in our sub-100nm technology nodes. We use a photoresist with low activation energy and sensitive PEB properties to achieve low K1 values, and thus eliminate the need of most optical proximity effect driven rules. We also adopt an OPC flow that detects over-varying shapes and performs OPC pattern library comparison and qualification in addition to the automatic incremental fix. These technologies simplify the DFM rules and minimize the number of the new problems coming from the traditional DFM method adopted by most foundries. We have achieved better than 90% for various products steadily.

Introduction

DFM (Design for Manufacturability) rules are typically restrictive design rules to strengthen shape integrity of on-chip conductors and semiconductors and enhance manufacturability and yield of semiconductor designs. We show some common DFM rules in Figure 1. As shown in Figure 1, successful work of Optical Proximity Correction (OPC) requires larger dimensions of spacing between conductors or conductors themselves than design rules for patterns that are vulnerable to optical proximity effects. Those restrictive rules reduce the effects and ensure the variation of conductor shapes is within a tolerance. For example, a larger spacing between two poly lines on the same active area (shown as S1 in Figure 1) reduces the relative variation of poly width because the aerial image now has weaker optical interference from the

Ex. 1016, Page 205

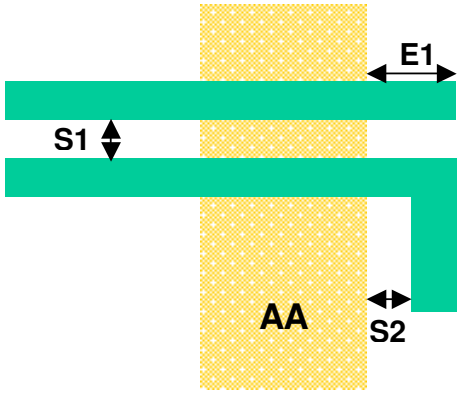


Figure 1. DFM rules require larger-than-design-rule dimensions for the spacing between two poly lines on the same AA (S1), poly extension on AA (E1), and spacing between an L-shaped poly line and AA (S2) on the same MOS.

neighboring poly. As a result of applying these rules, the chip performance is under controlled variation. Furthermore, the probability of random defects also drops in critical areas.

However, as the physical dimensions of semiconductor designs are pushed to sub-100 nm, ever-complicating design patterns require increasing numbers of DFM rules. Hundreds of new DFM rules have been released in the last decade among the advanced foundries and IDM's for sub-100nm technologies as compared to the 130nm node, according to our study. A systematic method of simplifying these rules effectively is needed. Furthermore, recent publication has also reported new manufacturing failures rising with the overly complicated and numerous DFM rules due to the interplay of different process effects that the rules missed in consideration [1][2].

In this paper, we present how using large process window patterning techniques and an OPC library qualification flow reduce the DFM rules required by the traditional method and still produce high yields for sub-100nm technologies. Because optical proximity effects are prevented using high throughput photolithography process, the amount of necessary restrictive design rules becomes much smaller. The new OPC flow further reduces the need by performing iterative OPC pattern library qualification through a hot-spot feedback loop. As a result, our sub-100nm technologies require a simplified set of DFM rules as compared to the other advanced semiconductor foundries, but maintain larger than 90% of yield in manufacturing various designs.

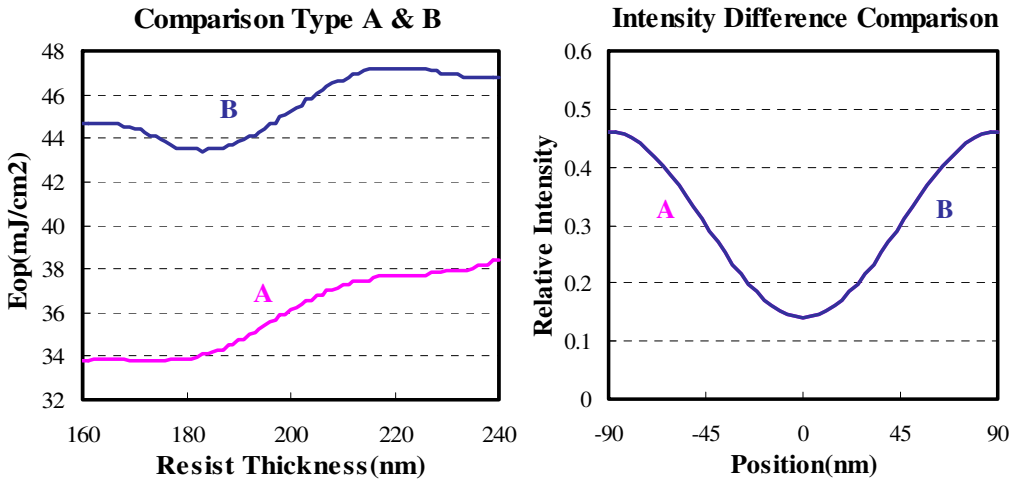
High Yield Patterning

We adopt photoresist of high photosensitivity and advanced materials to achieve high yield patterning. In Table I, we list the properties of the advanced photoresist (Type A) and compared to another one (Type B) in mainstream use of the industry. As shown in the table, Type A (36.12 mJ/cm^2) can sustain 19% less of operation energy (Eop) than

Ex. 1016, Page 206

Table I. Comparison of parameters between Type A and B photoresists.

		Type A	Type B
Scanner	Wavelength	193 nm	193 nm
	NA	0.80	0.80
	Sigma in	0.5	0.5
	Sigma out	0.8	0.8
	Pitch(equal L/S)	180 nm	180 nm
	Eop (mJ/cm2)	36.12	44.68
Resist	n/k	1.73/0.0	1.73/0.0149
	A/B/C	-0.05/2.31/0.065	0.0/1.50/0.0455



(a) (b)

Figure 2. (a) Operation energy Eop and (b) Relative intensity.

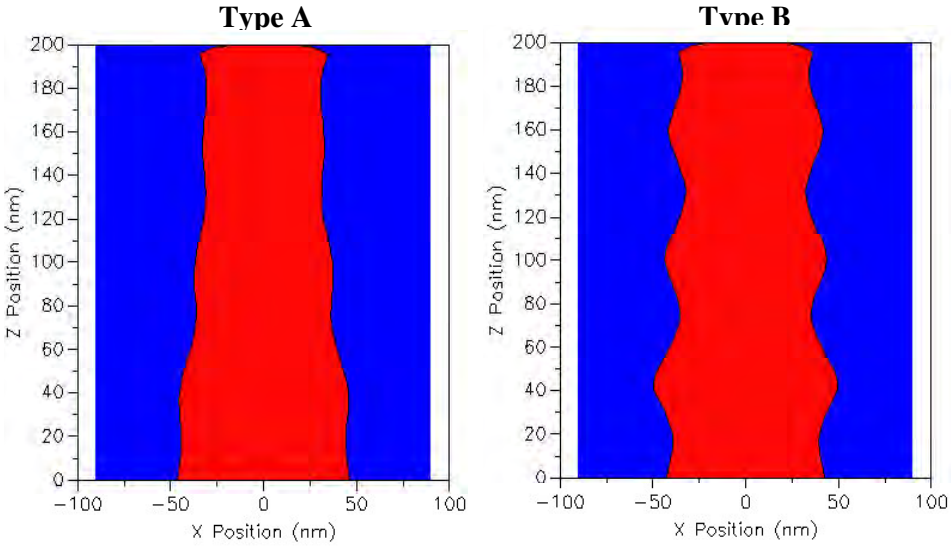


Figure 3. Vertical profiles.

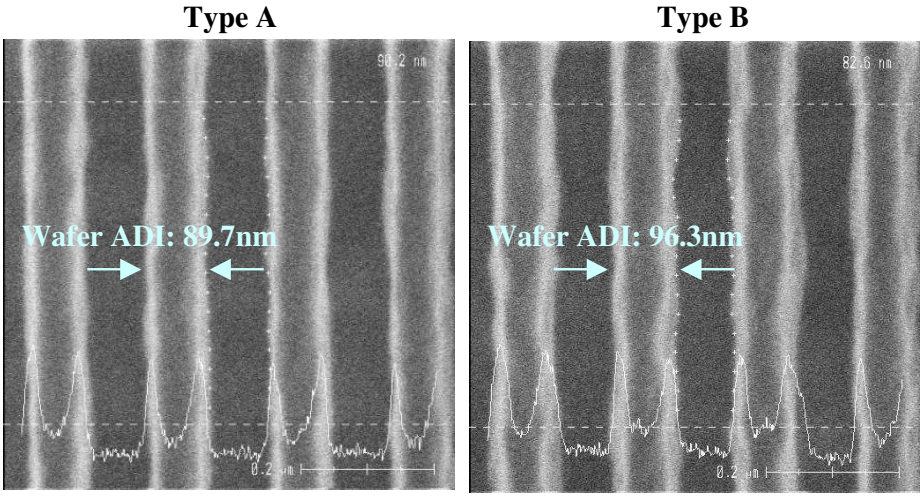


Figure 4. Type A and Type B ADI images on silicon substrate for 90nm line width.

Type B (44.68 mJ/cm^2) under the same scanner parameters because of low activation energy. The low Eop of Type A maintains throughout the thicknesses of photoresist, as shown in Figure 2 (a), and produces the same patterning intensities as Type B does, as shown in Figure 2 (b). Because of the same reason, as shown in Figure 3, the exposed vertical profile along the z direction of Type A has significantly (more than 50%) smaller standing-wave variation as compared to Type B. This small variation achieves minimum post-bake exposure (PEB) engineering and eliminates the need of multiple DFM rules to maintain low margins of critical dimension (CD) change.

We fabricated and compared after development inspection (ADI) silicon images of test patterns between the two types of photoresist on a 90nm process technology. The scanner parameters are shown in Table I and the machine power is 20W. The test patterns are parallel lines with both width and spacing equal to 90nm. We show the ADI images from both Type A and B in Figure 4. As shown in the figure, line width of Type A has a CD loss of 0.3nm, while Type B has a CD gain of 6.3nm. The average CD dimensions across the test patterns have approximately the same difference between Type A and B as well. With such a significant CD variation (7% of CD), Type B requires a larger spacing between lines to control the variations while Type A does not.

OPC Recipe Qualification Flow

We also use an advanced OPC flow to enhance the manufacturability and yield, as shown in Figure 5. In Figure 5, we compare the traditional OPC flow with ours. In additional to a hybrid rule-and-model-based OPC, OPC verification, and OPC data feedback (performed in the OPC verification step as well), a new OPC Recipe Qualification is performed in the advanced flow. Based on the variation detection and incremental fix performed by traditional OPC verification, OPC Recipe Qualification detects and separates severe shape or CD change violation spots (hot spots) and feed

Ex. 1016, Page 208

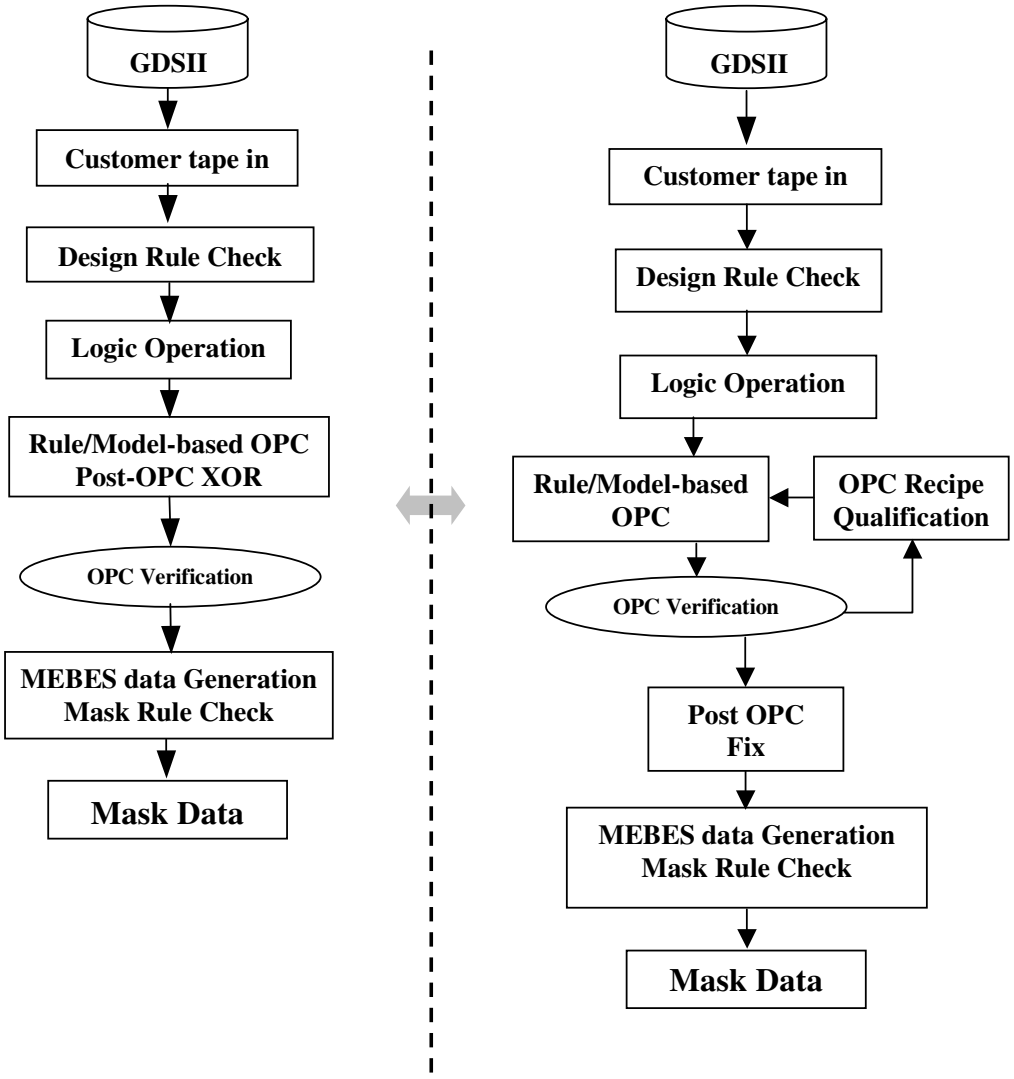


Figure 5. Comparison of traditional (left) and our (right) OPC flows.

back patterns of the hot spots to OPC pattern library and compare with the existing primitive patterns therein. After comparisons, new primitives to enhance OPC quality of patterns in the hot spots are added to the pattern library. Then OPC fix is performed automatically with this re-qualified and enhanced library. This detect-feedback-compare-qualify-enhance-fix iteration is performed until the number of OPC hot spots becomes zero. The direct effect of OPC Recipe Qualification is that the process window of OPC becomes large and adaptive to various designs. With this flow, OPC incurs much less lithographic variations, and therefore further reduces the need of DFM rules.

Ex. 1016, Page 209

High Yield with Simplified DFM Rules

With the above innovations, we have steadily achieved larger than 90% of yield in our sub-100nm technologies with a simplified set of DFM rules. The designs include logic, memory, and embedded systems. The number of DFM rules is more-than-50% less than the corresponding offering from other advanced foundries, as our study shows. In additional to OPC-driven DFM rules, we also offer rules for CMP effects and the interplay of process variations between OPC, CMP, and other processes as we discussed in Ref. [1].

Conclusions

The way for semiconductor foundries to achieve high yield by offering DFM rules needs re-innovation today. The reason is not only that immensely complicating DFM rules need a proper simplification, but also that some DFM rules have caused new manufacturing failures.

Our solution of simplifying DFM rules is presented in this paper. The approach is to fundamentally eliminate the need of restrictive design rules by improving patterning qualities. We adopt a photoresist with high photosensitivity to achieve low K1 values. As shown in silicon image, the CD variation of this new adoption is smaller than that of a mainstream photoresist recipe by one order of magnitude. This advancement reduces many DFM rules that are required otherwise. In addition, we also adopt an OPC flow that performs iterative OPC recipe qualifications to enlarge process window and ensure the shape integrity that may not be maintained by a traditional flow.

Therefore, we are able to use a simplified (more than 50% cut) set of DFM rules to achieve high yield patterning. We have achieved larger than 90% of yield in our sub-100nm technologies.

Acknowledgment

We thank Eric Guo for many meaningful discussions.

References

[1] Li-Fu Chang, " Collaboration of Semiconductor Foundry and Electronic Design Automation on Design for Manufacturability (DFM) in Sub-100nm Process Technologies", BASTS, October, 2007.

[2] X. Xu *et al*., "Modeling Process Impact on Cu/Low k Interconnect Performance and Reliability", University/Government/Industry Microelectronics Symposium, June, 2006.

ECS Transactions, 18 (1) 183-186 (2009)
10.1149/1.3096447 © The Electrochemical Society

Evolution of Wafer Inspection and Review Methodology

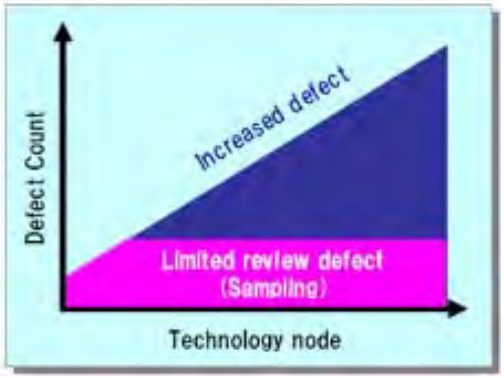
Ellis Chang, Allen Park

KLA-Tencor Corporation, Milpitas, California 95035, USA

Since the introduction of KLA 2000 series systems in mid 1980, Semiconductor manufacturing has widely adopted automated wafer inspection to detect defects during manufacturing to mitigate excursion and to reduce overall defect density. While the earlier emphasis in yield management was to detect the smallest defect possible, current environment calls for changes in both inspection and post processing technique that lead to identification of yield relevant defects in an efficient manner. Industry calls for intelligent way of inspecting the most critical areas on a device with high sensitivity and post inspection technique that utilizes advance techniques to generate defect pareto that highlights most important issues in defect data. Such methods are needed to meet both technical and financial goals of semiconductor companies.

OLD METHOD IN NEW ENVIRONMENT

Automated wafer inspection systems have been widely adopted in Semiconductor manufacturing over the past 30 years preceding the introduction of KLA 2000 series in 1984. In-line wafer inspection has significantly contributed to advance manufacturing technologies where defects in process can be detected early to reduce development time and to protect yield overtime. In earlier days, ability to detect defect was one of the main focus, but now there is a shift in requirement. Over the years, defect counts per wafer have exploded to as many as 1 million defects per wafer due to large wafer size and more sensitivity inspection technologies (Fig. 1). While overall detected count increased and critical defect size became smaller, typical defect review strategy has not change over the period, especially in the defect review area where random sampling is dominant. Such environment can produce defect pareto that is prevalent with nuisance defects (Fig. 2) [1]



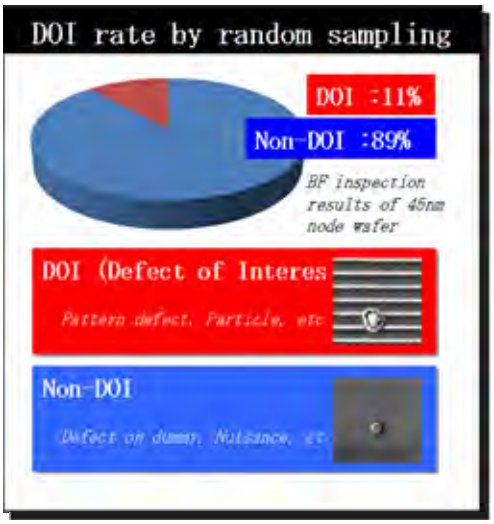


Figure 1 & 2. Defect review sampling remained constant as defect count increased. % Defect of Interest therefore decreased as a function of total defect count.

Ex. 1016, Page 211

TRENDS IN INSPECTION DEFECT MANAGEMENT

Traditionally, in-line monitoring strategy has focused mainly on random defects such as random particles. While detecting random particles are important, newer technology nodes exhibit systematic defects that are difficult to detect (Fig. 3) [2]. Even after detection, identification of such defect among high defect count presents a challenge where sample rate of 50 defects per wafer from 10^5 defects only represent 0.05% of data. With increase in systematic defect, more emphasis is taken for identifying the patterning issues early in the process development to reduce product introduction cycle. Some methods include use of Focus Exposure Matrix (FEM) or PWQ (Process Window Qual) are being used to identify systematic patterning issues. At the same token, marginal patterns discovered during device development are being asked to be monitored to detect failures that may result process variation and interactions. To achieve this goal, a method that uses part of design clip as library to monitor is implemented to effectively monitor critical pattern types (Fig. 4). In this method, design clips for each marginal pattern can be registered into a library and any occurrence of such pattern failures are effectively captured and classified.

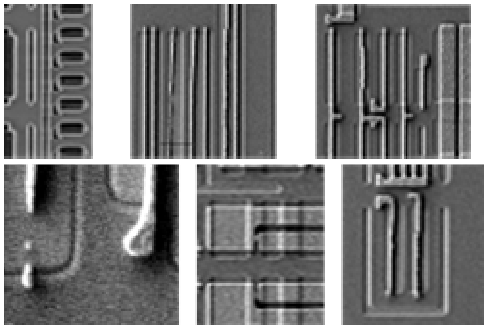


Figure 3. Example of systematic defects including bridging, necking, line thinning.

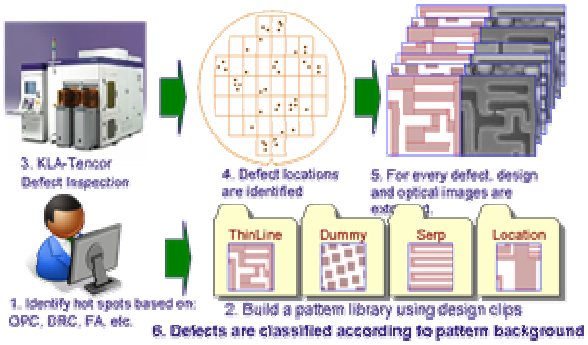


Figure 4. Method of monitoring marginal patterns in in-line monitoring.

APPROACH: PARADIGM SHIFT

To increase efficiency in detection and identification of critical defects, a new approach must be adopted. Reliance on simple defect filtering and size-based defect prioritization is not sufficient. To take optimal advantage of inspection and review budget, new intelligence must be employed on both inspection setup as well as review strategy (Fig. 5). Intelligence such as critical areas and hotspots resulting from design and simulations can be inserted into inspection strategy to optimize the available inspection capacity. During and upon completion of inspection, defect filtering based on multiple inspection attributes and binning allows users to quickly identify new defects while quantifying known defect types effectively. By taking advantage of knowledge in location of defect proximity in design space where design complexity is understood, yield relevance of each defect can be better evaluated to provide a prioritization of defects (Fig. 6). With design-aware inspection setup, now a more intelligent inspection recipe can be generated by focusing on areas in die that matter most. Such techniques have been successfully tested in both development and production Fabs. By introducing new methods, defect identification is more efficient and therefore process development time can be reduced by effectively analyzing experimental results (Fig. 7). With the ability monitor critical

Ex. 1016, Page 212

structures at ease, each set of experiment can be evaluated quickly to select the process conditions that yield best in printability. [3]

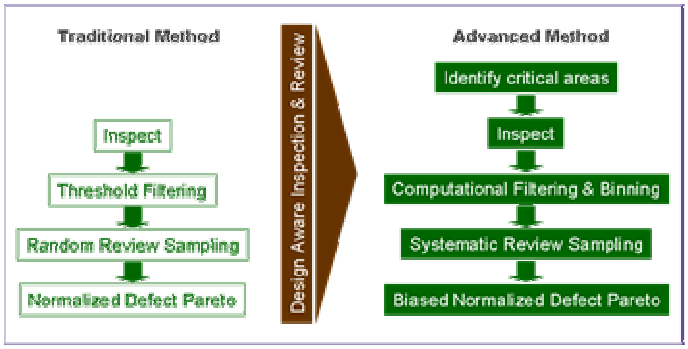


Figure 5. New method is needed to create operation excellence in chip manufacturing.

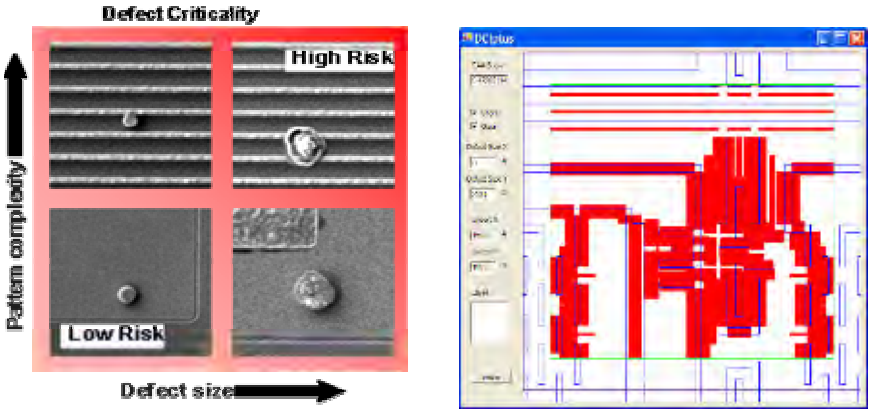


Figure 6. Defect criticality can be evaluated by combing defect attributes with design complexity to identify yield relevancy of each defect being detected in-line.

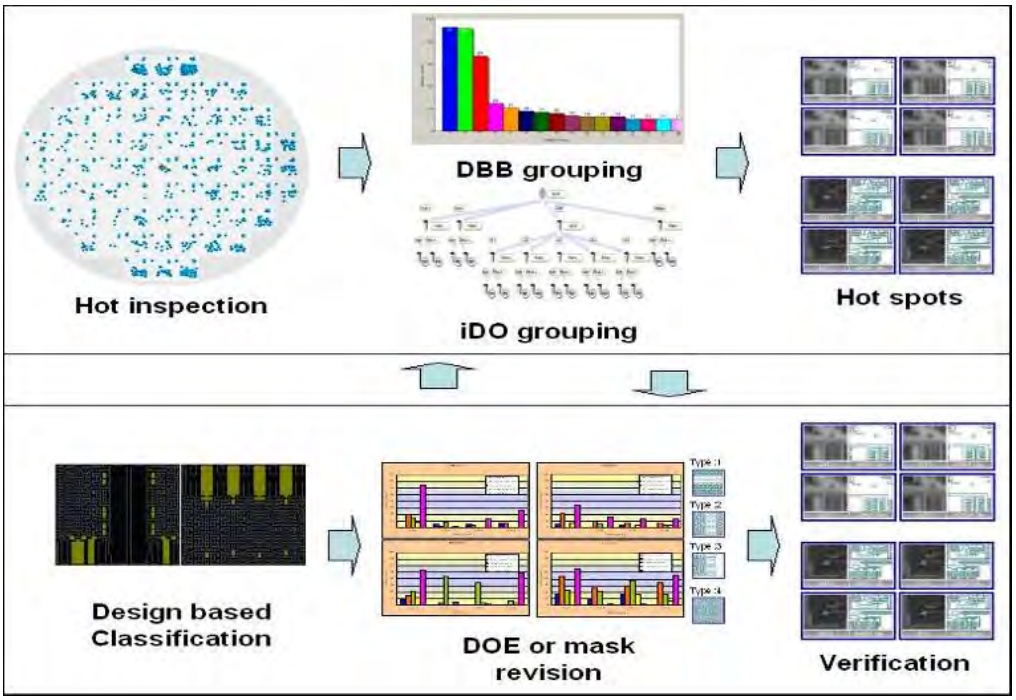


Figure 7. Integration of defect and design enables new product development and introduction faster with higher efficiency through reduced mask revision.

Ex. 1016, Page 213

CONCLUSION & DISCUSSION: SURVIVAL OF FITTEST

With new approach in in-line monitoring, defect issues can now be identified in a systematic approach. In an environment that requires both technology advantage and operation efficiency to stay competitive, a new method must be implemented to maximize the resources available for defect management. Defect inspection no long needs to be in a silo mode in Fab alone, but designer, litho and defect engineers must take advantage of newly available technology to identify random and pattern failures efficiently. Furthermore in-line inspection can now become design aware and be able to optimize the inspection budget to inspect where it matters most. As semiconductor advances to the next step in a competitive environment, integration of defect and design will be prevalent for operation excellence in defect management.

Acknowledgments

Authors would like to thank many individuals at K-T for useful discussions and development of new techniques. Authors would like to express appreciation to those customers at UMC, tsmc, IBM, Toshiba and number of other customers for validation of ideas and verification of outcome.

References

Complete citations should appear at the end of the text. Use the reference style that is shown below for all references.

1. Y. Sato, Y. Yamada, et al, *SPIE Advanced Litho*, 6922-37, (2008).
2. C. Huang, H. Liu, et al, ASMC, 1-1, (2008)
3. P. Jeng, C. Lin, et al, SPIE Taiwan, (2008)

Ex. 1016, Page 214

ECS Transactions, 18 (1) 187-192 (2009)
10.1149/1.3096448 © The Electrochemical Society

A Systematic Methodology For Low Yield Excursion Analysis

Xiao-Qing Xu, Chi Liang Lin, Thomas Tey, Michael C.C. Hsu

United Microelectronics Corporation (Singapore Branch), Singapore, 519528

Due to the complicated process and huge cost of a modern semiconductor wafer foundry, maintaining stable yield is very critical for competition. If the yield killer is not identified, no containment action can be taken; if there is no containment action, the impact will continue and the yield loss will enlarge. In this paper, we have proposed a systematic methodology for low yield excursion analysis. It integrates Kepner-Tregoe (KT) method and fishbone factor analysis to identify the yield killer in an efficient and effective manner. From the real case application, it is proven that KT method and fishbone factor analysis can be integrated and resolved low yield excursions successfully. This methodology is widely used in UMC Singapore Branch to handle FAB yield excursions of daily management and works very well.

Introduction

A semiconductor fabrication plant, which is usually abbreviated as FAB, is a factory where integrated circuits (IC) are manufactured. For a modern semiconductor FAB, which usually requires hundreds of steps to manufacture a wafer, low yield cases usually happened at the poorer process margin areas or due to any inline equipment drifts. Yield stability is very critical for every FAB to stay at competitive position. Hence, it is very important to find out the killer of the yield excursion quickly and correctly, and then the containment/corrective actions can be taken properly.

Traditionally, process integration engineers and product engineers are prone to use equipment commonality check and failure analysis (FA) to identify the killer [1]. However, commonality check sometimes lacks of correctness due to lots of common tools; FA has its limitation for long cycle time or can not proceed further if the failing site is unable to be located. Therefore, we need a methodology to find out the low yield's killer quickly and correctly in a more systematic way.

In this paper, we have proposed a systematic methodology to solve complicated FAB low yield excursion issues efficiently and effectively. This methodology not only combines most of low yield analysis methods that process integration and product engineers are using, but also provides a new approach which integrates Kepner-Tregoe (KT) method [2] and fishbone factor analysis [3].

KT method, a rational thinking process, is one of the most unique, documented analysis and decision-making methods, and is widely used for factories by both problem solving and decision-making [4]. It helps the decision maker to effectively collect status data, systematically organize and prioritize information, logically infer the possible cause and analyze impact. However, it is not easy for engineers to use the full-scale application of KT method directly to handle urgent cases due to tight schedule [5].

Fishbone factor analysis is another widely used analysis tool, which uses brainstorm to obtain all the possible killers and then uses 3 screening steps to find the possible killers.

Ex. 1016, Page 215

It's very helpful to use status analysis from KT method as an input to verify the possible killers from fishbone factor analysis. However, it has never been done in literature.

Therefore, in this paper, we propose a systematic methodology to combine KT method with fishbone factor analysis to find out the killer of yield excursion of a FAB in a more efficient and effective way.

Methodology

The methodology is shown in Figure 1. It starts from problem statement. Then background analysis helps us to understand the problem status and narrow down the killer to some specific suspected process loops. After background analysis, there are two analysis paths: One is commonality check, the other is 3W1H analysis (KT method) and fishbone method with 4 screening steps analysis. These two analysis paths will conclude to some suspected killers respectively. Finally, intersection check helps to identify the real killers from these two analysis approaches.

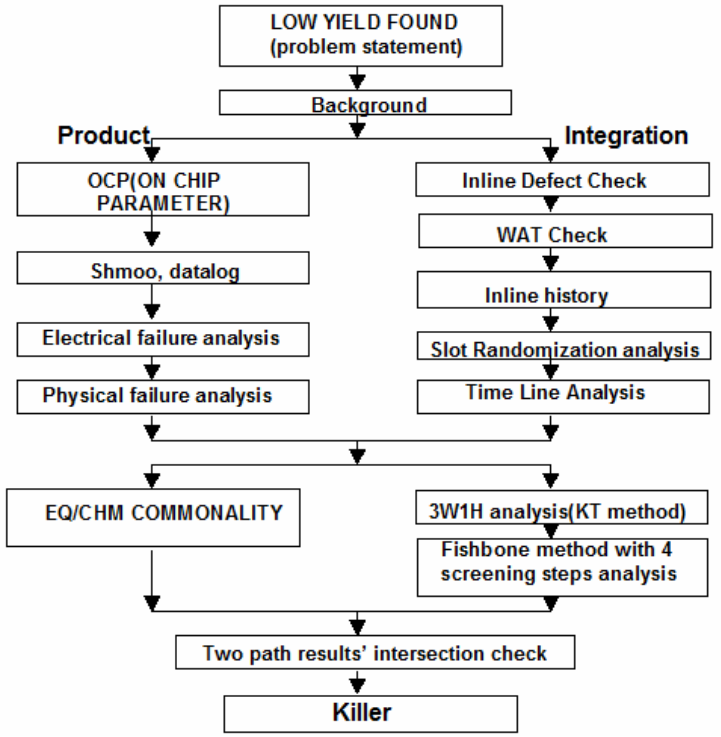


Figure 1. The systematic methodology for low yield excursion analysis

Problem Statement

The problem needs to be described clearly by a short statement including the object, what happen, deviation, impact, low yield wafer map and yield trend chart. The purpose of problem statement is to find "What's wrong" and to collect some basic information.

Background Analysis

At this step, we list out the facts and impacts of the problem. The facts describe the phenomena which mean what findings have you seen in different perspectives while the impacts describe the severity which means what results does the issue cause.

In this methodology, we suggest collecting the facts from two points of view for low yield excursions. One is from process integration engineer point of view and the other is from product engineer point of view. The former considers the facts including inline defect check, WAT (Wafer Acceptance Test) check, inline history (hold, rework, OOC, OOS, experiment taken, etc.), randomization slot analysis, and time line analysis, etc.

Ex. 1016, Page 216

The latter considers the facts including on-chip-parameter, shmoo (hard or soft fail), datalog (leakage testing items' information), electrical failure analysis (EFA) and physical failure analysis (PFA).

Background analysis is very important for killer identification of low yield issues. Some issues' killers can be identified immediately after effective background analysis. Even if the killer is not confirmed by background analysis, the suspected process can be narrowed down from the whole process to some specific process loop.

Commonality Check

Because most excursions will impact more than one lot, equipment commonality check is widely used. This paper provides a more advanced and systematic commonality check procedure. It consists of product commonality, equipment commonality and time commonality. The details of commonality check's procedure are shown in Figure 2.

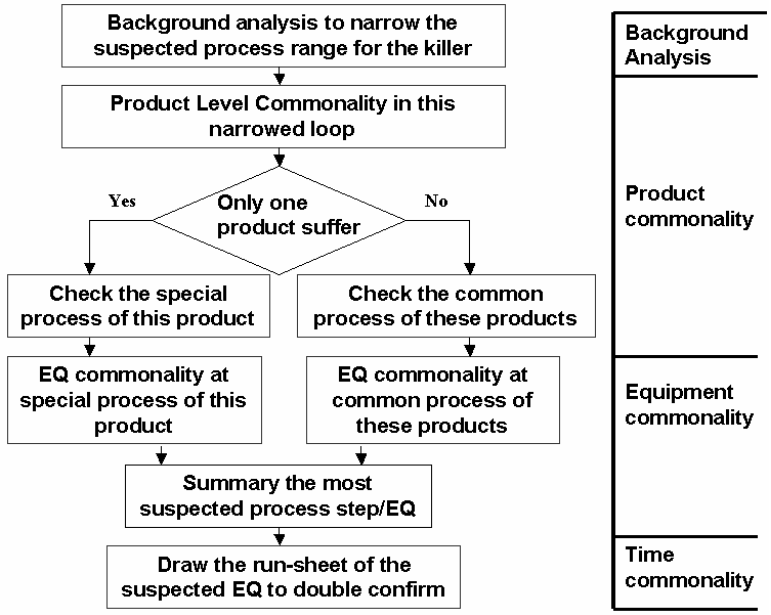


Figure 2. The systematic procedure of commonality check

3W1H Analysis (Kepner-Tregoe method)

In this paper, we use "3W1H" (What, Where, When and How many) and "Is/Is Not" in KT Method to collect the status of the problem. Based on 3W1H table, we can analyze the status of the problem by asking at least 11 questions which are object and deviation for " What"; geography and location for "Where"; when first, when reoccurred, continuous or not, which step for "When"; how many objects, how big, and trend for "How many". "Is" means it happens on a certain thing or event while "Is Not" means a similar thing or event should happen but not happen with respect to each question in 3W1H table. Then a critical clue can be chosen from 3W1H table based on the fundamental theory, practical experience and professional knowledge about the issue.

Fishbone Method with 4 Screening Steps Analysis

As shown in Figure 3, the chosen critical clue is used as the input of this step. Based on the critical clue, we can draw a fishbone chart to list all possible killers. Then fishbone method with 4 screening steps analysis can help us to filter the possible killers. The method is described as follows:

Step 1: Screen out the possible killer if it changes but does not result fish head to change. The screen out rule is: if the result is "Yes", it is marked by a "Circle" and keep

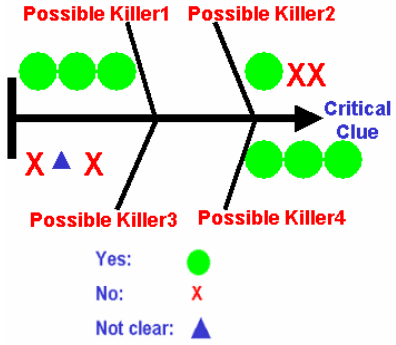
Ex. 1016, Page 217

it temporarily; if the result is "No", it is marked by a "Cross" and is screened out; if the result is "Not clear", it is marked by a "Triangle" which means it needs more data to verify and keep it temporarily.

Step 2: Screen out the possible killer if it cannot be changed or if it did not change recently. The screen out rule is the same as Step 1.

Step 3: Screen out the possible killer if the fish head changes but does not result the possible killer to change. The screen out rule is the same as Step 1.

Step 4: Use 3W1H table to check the explanation ability for remaining possible killers and choose the most explainable one. The screen out rule is: for each question in 3W1H table, if the result is "Yes", it is marked by a "Circle"; if the result is "No", it is marked by a "Cross"; if the result is "Not clear" it is marked by a "Triangle". The one with less number of "Circle" will be screened out. That is, the one with the most number of "Circle" is the most probable killer.



		IS	IS NOT	Possible Killer1	Possible Killer4
WHAT	Object			●	▲
	Deviation			●	●
WHERE	Geography			▲	▲
	Location			●	●
WHEN	When First			●	●
	When Reoccured			●	●
	Continue or not			▲	●
	Which Step in the Process			●	▲
HOW MANY	How Many Objects			●	x
	How big			▲	x
	Trend			▲	▲
				8 ●	5 ●

Figure 3. Fishbone method with 4 screening steps analysis

Intersection Check

After commonality check and fishbone method with 4 screening steps analysis, there may be still several suspected killers. Then intersection check can be implemented to narrow down the real killer if they have intersection. However, if there is no intersection, we should go back to cross check the commonality and fishbone method with 4 screening steps analysis or even redo both of them with these useful feedbacks.

Case Illustration

In this section, we would like to follow the systematic methodology mentioned above to solve a real case of yield excursion in a FAB.

Problem Statement

The problem is: Product 1/2/3 (Object) yield trends down (What happen) 8% (Deviation), impacts 8 lots/200 wafers (Result), and low yield wafer map is random.

Background Analysis

We check the items listed in background in two perspectives. From product engineer's point of view, we find major bitmap is single bit failure by checking EFA. From process integration engineer's point of view, we check the timeline of the process from wafer start to wafer ship out as shown in Figure 4. The green lines are good lots, while the red lines are bad lots. Timeline's X-axial is the process step and Y-axial is the process calendar time. From step16 to step116 during the period of 5th Sep to 13th Sep, all bad lots are continuous and no good lot is in this period. Therefore, this time period from step 16 to step 116 is highly suspected.

Ex. 1016, Page 218

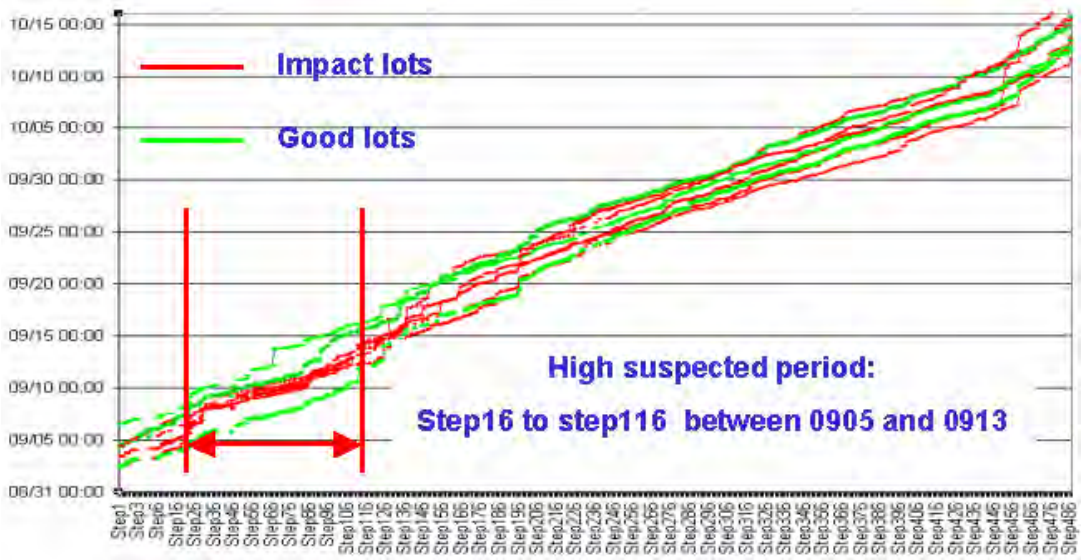


Figure 4. Timeline analysis

Commonality Check

Doing commonality in suspected process loop to find thick gate oxide (TG) etch clean tool C-03 is the common tool. Then, we check C-03's run sheet, as shown in Table I, and find all impacted lots are at the same chemical change cycle of C-03.

TABLE I. C-03 run sheet

Product	Lot No.	Equipment	Move in time	Yield	Impact
2	A	C-03	9/4	96.3%	Baseline lot
3	C	C-03	9/4	95.1%	Baseline lot
C-03 Chemical Change					
3	B	C-03	9/5	**87.4%**	**Impact lot**
1	D	C-03	9/5	**85.6%**	**Impact lot**
3	E	C-03	9/5	**82.5%**	**Impact lot**
1	G	C-03	9/5	**89.3%**	**Impact lot**
1	F	C-03	9/5	**87.0%**	**Impact lot**
2	K	C-03	9/6	**86.5%**	**Impact lot**
1	I	C-03	9/6	**88.2%**	**Impact lot**
3	H	C-03	9/6	**86.9%**	**Impact lot**
C-03 Chemical Change					
2	J	C-03	9/7	96.3%	Baseline lot
1	L	C-03	9/7	94.5%	Baseline lot

3W1H Analysis (Kepner-Tregoe method)

Next, we use 3W1H table with 11 questions of what, where, when and how many with respect to "Is/Is Not" to collect the status of the problem as shown in Table II . From 3W1H table we choose SRAM single bit failure as the critical clue.

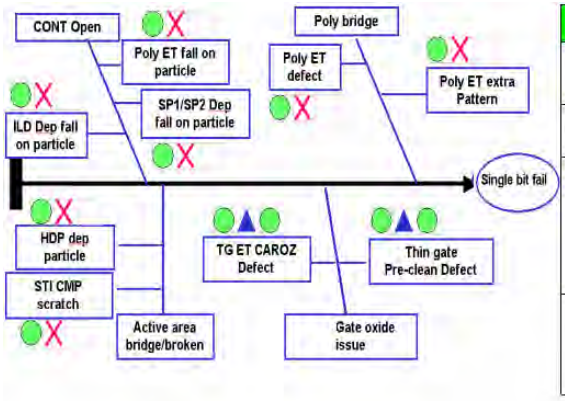
TABLE II. Demonstration of 3W1H analysis (Kepner-Tregoe method)

		IS	Is not
What	Object	Product 1,2,3	Product 4
	Deviation	**Bin1 high, single bit fail**	IDDQ
Where	Geography	Every wafer in the lot	First 3 wafers in the lot
	Location	Random	Wafer Right
When	When first	200x/09/05	Early or later
	When reoccurred	200x/09/06	Early or later
	Continue or not	Continue	Every 3 days
	Which step	CP Testing	WAT Testing
How Many	How many objects	8 lots	More or Less
	How big	Yield Loss 4~12%	More or less
	Trend	Trend down	Trend up

Ex. 1016, Page 219

Fishbone Method With 4 Screening Steps Analysis

We use the critical clue of SRAM single bit fail as the fish head, and draw the fishbone chart as shown in Figure 5. Then we apply the first three steps in fishbone 4 screening steps analysis to screen out false suspected killers. The suspected killer which has at least one "Cross" will be screened out. Then, the suspected killers left are: TG etch clean and thin gate pre-clean. For these two suspected killers, we apply the 4th step of fishbone method with 4 screening steps analysis to identify the most probable killer which is TG etch clean since it has the most numbers of "Circle" as shown in Figure 5.



		IS	IS NOT	TG ET Clean	Thin gate pre-clean
WHAT	Object	Product 1,2,3	Product D	●	▲
	Deviation	Natural good bin worse,single bit	IDDQ	●	●
WHERE	Geography	Every wafer/lot	3 wafers/lot	▲	×
	Location	Random	Wafer right	●	●
WHEN	When First	200x/09/05	Early or later	●	●
	When Reoccured	200x/09/06	Early or later	●	●
	Continue or not	Continue	Every 3 day	▲	●
	Which Step in the Process	CP Testing	WAT Testing	●	▲
HOW MANY	How Many Objects	8lots	More or less	●	×
	How big	CP yield loss 4~12%	More or less	●	▲
	Trend	Trend Down	Trend up	▲	▲
				8●	5●

Figure 5. Demonstration of fishbone method with 4 screening steps analysis

Intersection Check

In this case, both commonality check and fishbone with 4 screening steps analysis have pointed out the most probable killer is TG etch clean process, which hints us that it is the real killer. Then we can take containment/corrective action to stop the impact.

Conclusion

In this paper, we have proposed a systematic methodology to integrate KT method and fishbone factor analysis for FAB yield excursion analysis in an efficient and effective manner. A real case of FAB low yield is also demonstrated to prove the effectiveness of the methodology. This methodology is widely used in UMC Singapore Branch to handle FAB yield excursions of daily management and works very well.

References

1. X.Y. Pao and S.Q. Liu, *Practice of CurrentTtransmission Electron Microscope*, Tsang Hai Book Publishing Co. (2008).
2. C.H. Kepner, and B.B. Tregoe, *The New Rational Manager*, Princeton Research Press, NJ (1981).
3. M. Suyama, *Concept of Quality Control and Problem Solving Techniques*, Komatsu career creation LTD, p.19
4. E. Lumsdaine, and M. Lumsdaine, in *Creative Problem Solving*, IEEE Potentials, 13(5), 4-9 (1994).
5. C.L. Lin, C. Wang, J.F. Chen and C.C. Hsu, in *Development and Application of a Systematic Problem Solving Methodology — Using Semiconductor Wafer Foundry as a Case Study*, Journal of Quality Vol.15, No.5, Taiwan (2008)

Ex. 1016, Page 220

ECS Transactions, 18 (1) 193-197 (2009)
10.1149/1.3096449 © The Electrochemical Society

Use FPGA Configuration Memory For Backend Defect Capture

G. Qian[a], J. J. Shi[b]
[a,b] Xilinx Asia Pacific, 5 Changi Business Park Vista, Singapore 486040

Abstract

In deep sub-micro era, with the increase of circuit complexity, more and more metal layers are used for inter-connection. Chance for killing defects coming from backend metal layers is significantly increased. Especially for lower metal layers with smaller features, defects on them may be difficult detected by conventional current-based isolation techniques. Therefore development of defect monitoring method becomes critical for wafer fabrication yield improvement.
SRAM-based FPGA is already well known for front-end defect capture capability. Studies also demonstrate its capability to capture defects on higher interconnection layers with specially designed test pattern. [1] With the dense and segmented data-line structure, the configuration memory can be used as a much simple approach to monitor defects on lower metal layer
In this paper, we will go through the architecture of a 90nm FPGA and its configuration SRAM. By study the electrical data-line failure signature, we will discuss the corresponding isolation techniques. Finally we will show some defects captured by this methodology.

FPGA architecture

In SRAM based FPGA, the memory cells are used to control the on/off state of the corresponding logic block like Look Up Table, MUX, Flip-Flop and their internal functions. Organized in the matrix similar to the pure memory array, cells are placed next to each other along the word line direction. While along the bit line, various logic resources are placed in-between different frames. Only one of the data-line/data-line-bar is used for read/write operation, while the other is connected to ground for fast reset purpose.

Data lines are running on metal-3 across the whole chip, together with other routing signals or power/ground lines with minimum design rule for most chip area. Given the width of metal-3 signal line is 0.14um, assume the minimum defect has the same size, a rough estimation show its coverage is around 32.4% of the core logic region.

The FPGA can be separated into many repeating logic units called CLB, short for Combination Logic Block. To ensure successful transmission of configuration data to the CLB at very end of die, data repeaters are inserted along each data line for distance of every two CLBs. In other words, the maximum un-disrupted Metal-3 data line length is 2 CLBs. During FPGA configuration, the bit stream data reach individual memory cells through the corresponding data line. The data also can be retreated by the reversed

Ex. 1016, Page 221

operation called read back. Data repeater also includes pre-charge circuit, which will pre-charge the data line segment to Vcc, before read operation.

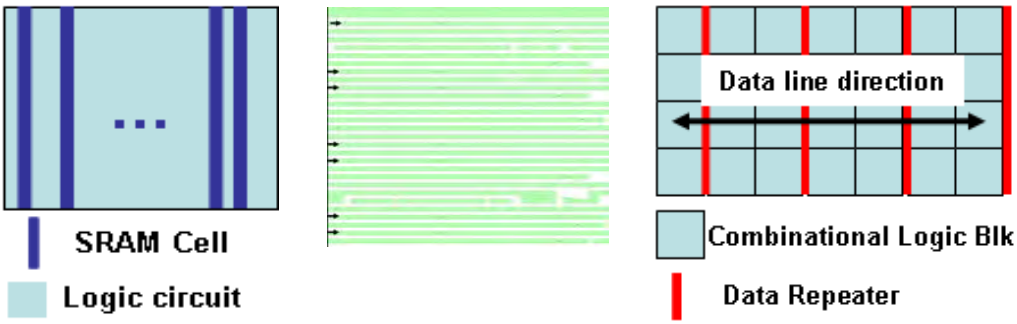


Figure 1. FPGA configuration memory architecture

FA methodology

The dataline failure we will discuss is conventionally described as column failure in SRAM. Consider the similarity between SRAM and FPGA, many of the analysis works done on SRAM can be applied here.[2] Test patterns used for FPGA configuration memory testing are 0's; 1's; checkerboard and reverse checkerboard, same as the SRAM. For test-time/cost saving purpose, only first 10 failures are recorded by production test program. Full read back data of the whole chip is available by engineering test program. Collected data log will go through script processing to generate the necessary address which can be used to identify the failure location. Below we will discuss on the possible open/short of data line and the corresponding electrical signature.

Assume an open on the data line:

Memory cells along this data line, between read/write port and the open fault should operate as normal, cells beyond the open will neither receiving nor sending any bit signals. When those affected cells are selected for read, only pre-charged value can be retreated, which means *ONE* is the readout for all cells beyond the open fault. Considering the data line is actually bit line bar, affected cells will fail 0's, and alternative failure on checkerboard and reverse checkerboard pattern. Base on the translated full die read back data, fault location can be easily located between the first failure cell and last good cell. (Figure 2)

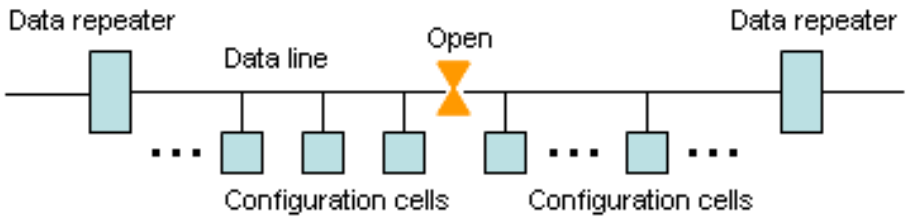


Figure 2. Illustration for an open failure along dataline

Assume a short between data line and its neighbor:

The data line will be set to the same voltage potential as its shorted neighbor. If the neighbor metal line carries high signal or power supply, and is strong enough to beat the write driver from data repeater, all memory cells beyond and include the affected segment will only receive ONE on their bit line during the write operation. Subsequently only high will appears on the data line for those affected cells during read operation. Therefore affected cells will fail 1's and alternative of checkerboard and reverse checkerboard patterns. Opposite conclusion can be drawn when neighbor metal line carries ground signal. Base on the translated full die read back failure addresses, fault location can be traced within the first segment along the affected data line. (Figure 3)

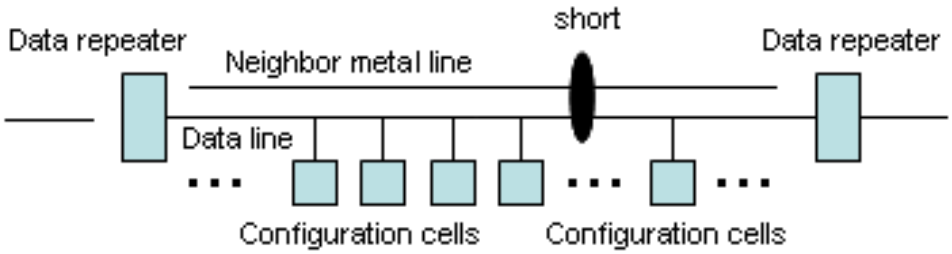


Figure 3. Illustration for an short failure along dataline

FA finding

Following the principle discussed in the previous section, here we will show some actual PFA finding.

PFA finding for case of short

Case-1: Read back data show this die has single data line failure starting from the beginning of one segment and failure pattern is 1's. Base on the analysis in previous section, the failure is most likely to be a short along that particular data line within the first failure segment (2 CLBs). Below is the short defect found by SEM at the isolated location. (Figure 4)

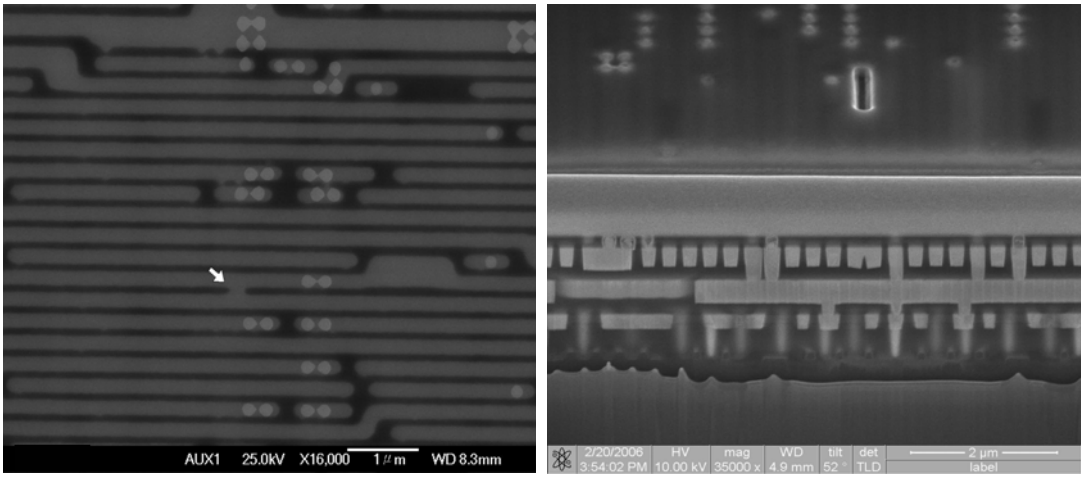


Figure 4. Short failure observed at the isolated location, right is the FIB cut view

Case-2: This is another similar case with single data line failure starting from the beginning of one segment, but fail 0's pattern. Following the same approach, the expected short should be located within the first failure segment (2 CLBs). Since the failure pattern is 0's, the short can be further narrow down to region with power supply nearby, assuming all nearby signals are reset to ground. Below is the final SEM photo of the found defect. (Figure 5)

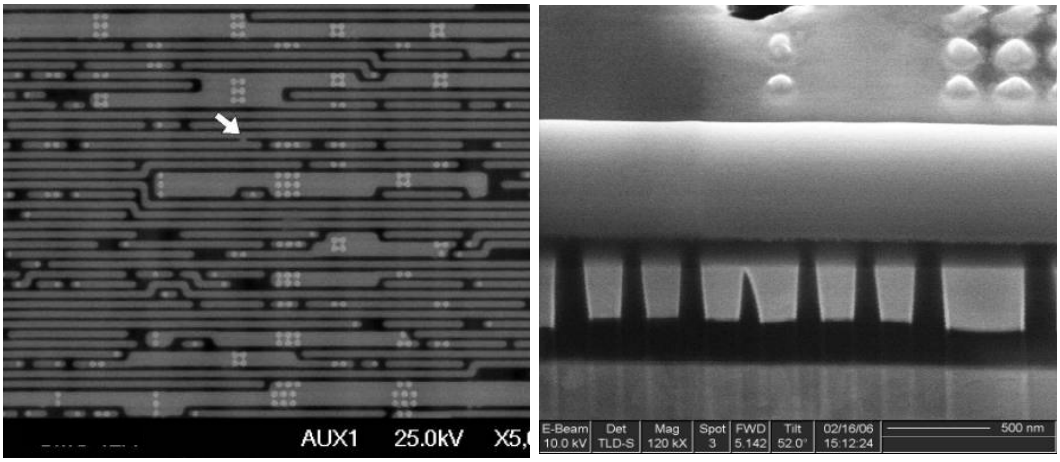


Figure 5. Short failure observed at the isolated location, right is the FIB cut view

PFA finding for Open Failure

Case-3: different from the above two cases, this unit show single data line failure starting from address within one segment, failure pattern is 1's. as discussed in previous section, this is a quite typical open failure signature. Memory cells beyond the open defect are no longer accessible in read/write operation. The data received during read operation is just the pre-charge value, which is 0's for corresponding pattern.

The open fault should located between the first failure cell and its nearby passing cell along the data line. Below is the SEM photo of the open defect found exactly at the suspected location. (Figure 6)

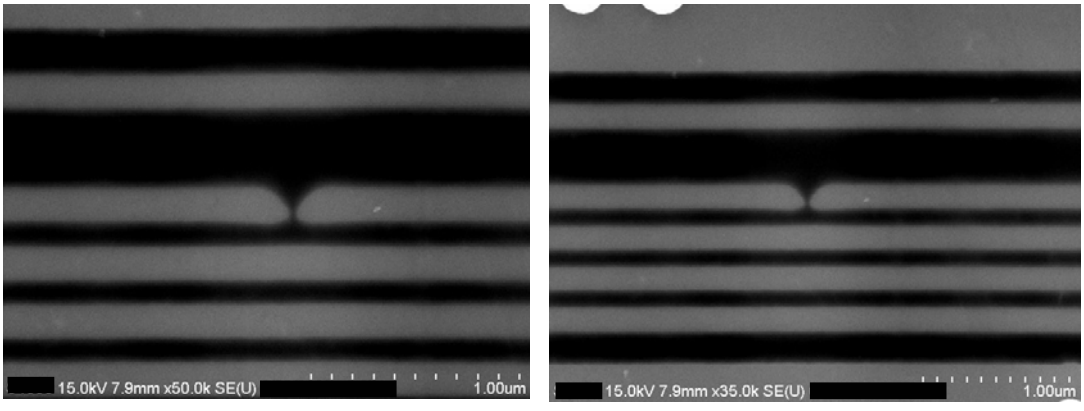


Figure 6. Open failure observed at the isolated location

Conclusion

FPGA configuration memory data lines, with width of minimum design rule, provide more than 30% coverage of metal-3 layer, it is ideal for defect capturing. Base on the failure address translated from the test result, defect type (open/short) can be predicted and the fault location also can be narrow down to manageable inspection region. Real physical failure analysis cases show the approach is valid and accurate.

References

1. Jenny Fan and Dave Mark *International Symposium of Test and Failure Analysis* Page(s): 409 – 413 (2004)
2. Z. G. Song, G. Qian, J. Y. Dai et.al *International Symposium of Test and Failure Analysis* Page(s):103 - 106 (2001)

Ex. 1016, Page 225

ECS Transactions, 18 (1) 199-204 (2009)
10.1149/1.3096450 © The Electrochemical Society

Functionality, Yield and Reliability Analysis of SiGe Micro-mirrors using Automated Optical Measurement Techniques

I. De Wolf[a,b], J. De Coster[a], O. Varela Pedreira[a], L. Haspeslagh[a], and A. Witvrouw[a]

[a] IMEC, Leuven, Belgium
[b] Department MTM, K.U. Leuven, Belgium

This paper discusses optical systems for automatic, wafer or chip level measurements of metrology, functional yield and reliability of MEMS. The functionality of the systems is demonstrated on megapixel arrays of SiGe micro-mirrors. The back-bone of the systems is formed by existing commercial systems, but home-developed procedures and software were written to extend their functionality and applicability.

Introduction

Micro-Electro-Mechanical Systems (MEMS) can be described as small mechanical machines (nm to mm scale) such as switches, resonators, mirrors, accelerometers, pressure sensors, actuators etc. They can be single devices, or grouped in arrays of thousands of devices working in unison. They have one thing in common which distinguishes them from CMOS devices: they can move. Because of this motion, problems such as stiction, fatigue, creep, thermal or stress induced deformation, etc. come into the picture and have to be studied. In many cases, this requires optical instrumentation. There exist many optical systems that can be used for this purpose, however, most of them are not adapted for automatic, wafer level measurements which allow to check in a fast, preferably automatic way the uniformity of tilt, buckling, stand-off etc of the devices within a chip or over the complete wafer. There is also a need for optical systems that can detect defective devices (stuck, damaged, etc), systems that can monitor variations in performance such as differences in motional range or resonance frequency, and systems that can monitor the reliability of a large number of MEMS.

In this paper we discuss three systems that are based on optical measurement principles and that can be used for functionality, yield, metrology and reliability testing of MEMS. The back-bone of the three systems is formed by existing commercial systems: a Differential Interference Contrast (DIC) microscope, a vertical scanning interference system and a scanning laser Doppler vibrometer. The functionality of the systems is demonstrated on a monolithically integrated 10cm² 11 megapixel SiGe micro-mirror array that was developed at IMEC for high-end industrial applications (1,2).

Samples

7.7µm x 7.7µm SiGe micro-mirrors were fabricated on top of 0.18 µm HV-analog-CMOS wafers (1, 2). The mirrors are actuated through SiGe electrodes, as shown in Fig. 1a. Fig. 1b shows schematically the array of mirrors. Fig. 1c shows a SEM picture of part of the array. The mirrors are connected to the underlying CMOS through SiGe electrodes and W-vias. Approximately 35M W-vias are used to connect the SiGe micro-mirror array consisting of 11 MPixels. In addition to these megapixel arrays, also test samples consisting of 24 mirrors in an array were used to test the systems.

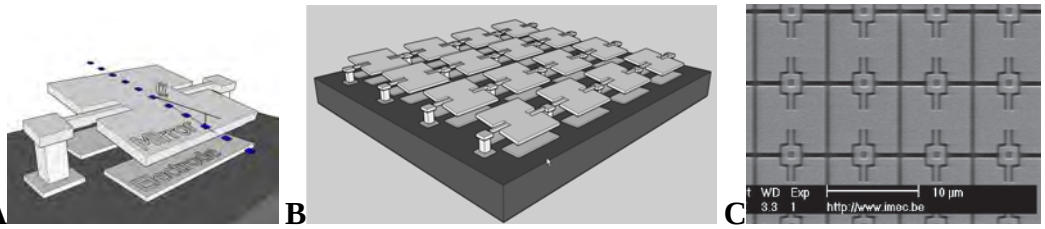


Figure 1. A: Schematic of 1 micro-mirror. B. Schematic showing several mirrors in an array. C. SEM figure of part of the array of micro-mirrors.

Optical measurement techniques

Profilometry

During processing development of MEMS, it is important to check the profile of the movable parts. In the case of micro-mirrors, for example, the movable mirror should in general be as flat as possible. In general, such measurements are performed through vertical scanning optical interferometry, which is at this moment one of the most important enabling measurement technologies for MEMS development (3). It has the advantage of providing the required information in a fast, fairly simple way.

The challenging part for the SiGe array application, was that a large surface ($10cm^2$), containing a huge number of identical very small MEMS (11 megapixels) had to be measured fast and accurately. For this purpose, a conventional profilometry system (Wyko NT3300D) was equipped with an X-Y stage. The best compromise between a good resolution and high number of MEMS in one field-of-view (FoV) was obtained with the 50x objective. One FoV contained 165 micro-mirrors. In order to cover the whole area of the array, 66181 non-overlapping FoVs had to be measured and stitched in a perfect manner. An additional challenge was that any drift of the system that would result in an incorrect stitching of the FoVs had to be compensated for. Software and methodology were developed for the two main functions of the system: an analysis module and a measurement module (Fig. 2), which work in parallel to increase the throughput of the system (4). With the resulting configuration, measurement and data analysis take 3.91 seconds for each FoV.

For the analysis module, first the data are pre-processed to correct for instrumentation related issues such as tilt of the sample or measurement head. This tilt is removed using a least-square fitting technique. Next the FoV is segmented into single MEMS. This is done using a set of templates of the devices and correlating this with the acquired FoV image. A template can be a photograph taken from the MEMS, or even a design lay-out. Next each single device is analyzed and the desired parameters are extracted. In the case of the micro-mirrors, this is done by fitting a parabolic surface to the data. From this fit, the cupping, tilt and stand-off of each mirror are deduced.

For the measurement module, the sample is moved using the automatic X-Y stage and at each position a FoV is collected. The main problem here is stage drift, causing an incorrect assignment of measurement results to measurement positions. To avoid this problem, the stage movement was corrected for any drift using a feedback algorithm. As a result, each FoV is correctly assigned to the position where the measurement is done, and many FoVs can be stitched together to get an image of the whole region of interest, whether they are overlapping or not.

Ex. 1016, Page 228

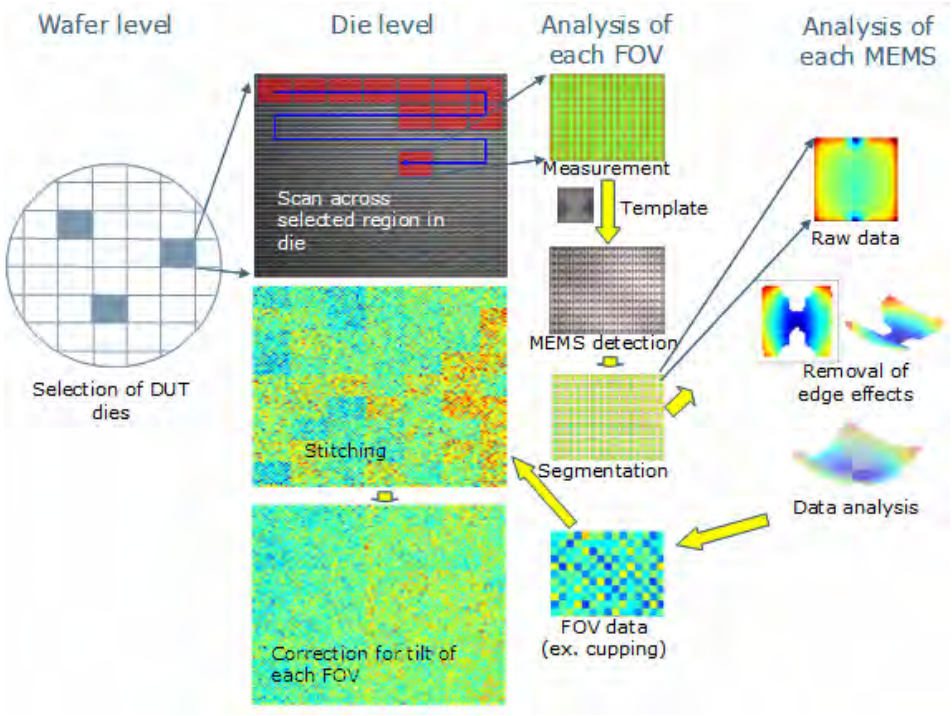


Figure 2. Profilometry measurement procedure.

Fig. 2 shows an example of a cupping map obtained from 63 FoVs (die level, bottom figure). An example of a cupping map result obtained from 16262 FoVs providing info on an array of 2.66 million devices is shown in Fig. 3. These results, obtained on a test chip, helped to further optimize the SiGe micro-mirror process. Also statistical analysis of these data is provided (Fig. 3 right). Such measurements can be done within one array, within one chip, or on wafer level.

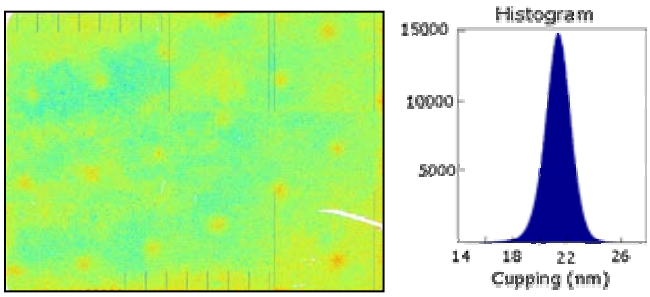


Figure 3. Left: Profilometry test results showing the cupping of 2.6 M devices on a test chip. Right: Histogram showing the cupping statisctical variation

Differential Interference Contrast (DIC)

A microscope with DIC-optics is used for functional inspection of the micro-mirror arrays (5). When a bias voltage is applied, the mirrors tilt by a certain angle and as a consequence their intensity in the DIC image changes. The purpose of the inspection

setup is to detect devices with inappropriate deformation by measuring the intensity of each device while sweeping the bias voltage. In other words, the tilt-versus-bias relationship of the mirrors is measured in an indirect manner using intensity as a metric of functionality.

The setup comprises both standard and custom hardware. The base component is a semi-automatic wafer prober, which has been equipped with a DIC microscope. In order to enhance the throughput of the system, the magnification of the microscope is chosen such that a maximum of devices can be measured in one FOV, while still obtaining enough data points for each device (i.e. pixel size of the camera ~ resolving power of the lens). A custom probecard with well over 200 pins allows interfacing the CMOS circuitry below the micromirror array with a series of dedicated electronics boards that implement analog as well as digital communication with the chip. The whole system is automated using Labview software. The use of mirror intensity as a measure of device functionality was validated using a laser Doppler vibrometer (LDV). The results of this experiment are shown in Fig. 4, left.

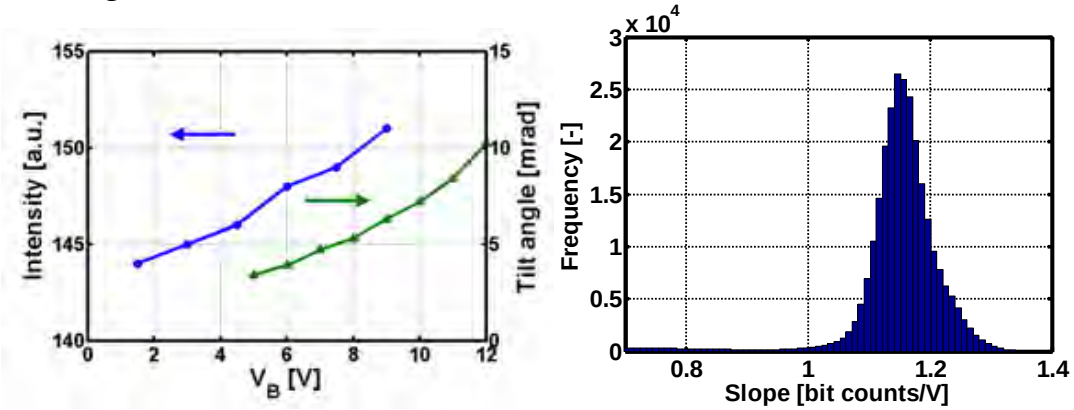


Figure 4: Left: Measured DIC intensity (left axis) and measured tilt angle (right axis) using LDV as a function of the bias voltage V_B. Right: Histogram of the slope that was extracted from the measured DIC-intensity-versus-bias-voltage characteristic.

The contrast between actuated and unactuated states of a device is hence used as an indirect measure for the amount of tilt and thus for the functionality of the device. The functional yield test therefore essentially consists of evaluating the intensity of all devices at various levels of the bias voltage V_B. A full yield inspection is performed by moving the microscope in a zigzag pattern above the wafer, and capturing images at each location as illustrated in Fig. 5.

This procedure results in histograms like the one shown in Fig. 4, right. The horizontal axis in this figure shows the slope that was extracted from the measured intensity during the bias voltage sweep. Several factors contribute towards the spread of this histogram; the main contribution is of course the physical distribution of device parameters within the chip, but several other factors play a role as well: the non-uniformity of the microscope illumination, the drift of the setup over the duration of the measurement, mechanical vibrations, non-flatness of the chip causing de-focusing of the images etc. all give rise to uncertainty on the measurement result. The influence of illumination is corrected for using a calibration procedure where a paraboloid is fitted to a series of captured images (e.g. 100 images at random locations on the wafer); the average of these surfaces is calculated and the mean value of the surface is shifted to 0. During subsequent measurements, this average surface is used as a reference and is subtracted from acquired images.

Ex. 1016, Page 230

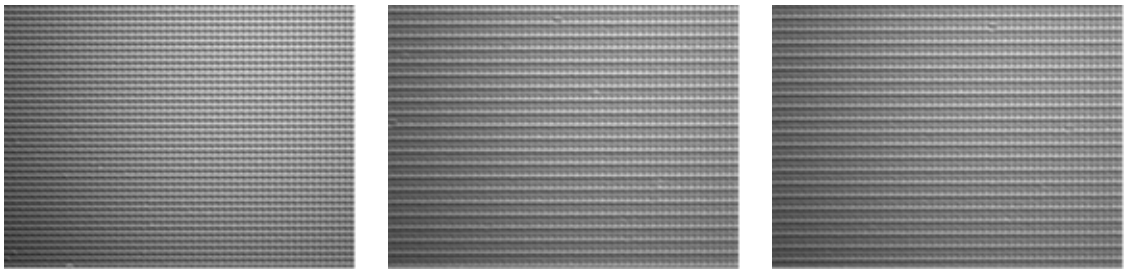


Figure 5: Left: image of unactuated device array. Middle: image of device array where rows are alternately tilting left and right. Right: image of device array tilted in the opposite sense as in the middle picture.

As mentioned above, the intensities that are obtained from the calibrated and segmented images are stored in a database. This database allows the distribution of defects over the wafer surface to be quickly visualized and correlated with other defect inspections that might have been done previously (e.g. during processing).

Laser Doppler Vibrometry

The third technique is making use of a scanning laser Doppler vibrometer (LDV) to monitor the motion of the mirrors with < 50 pm resolution (4,7,8). The system measures the motion in a single point, but by scanning the laser beam across the sample one can measure a pre-defined number of points, covering several MEMS. Customized software was written, allowing the system to automatically extract time-dependent information such as vertical motion, time dependent tilt, resonance frequency, step response etc. (5). The procedures followed for such tests are demonstrated in Fig. 4.

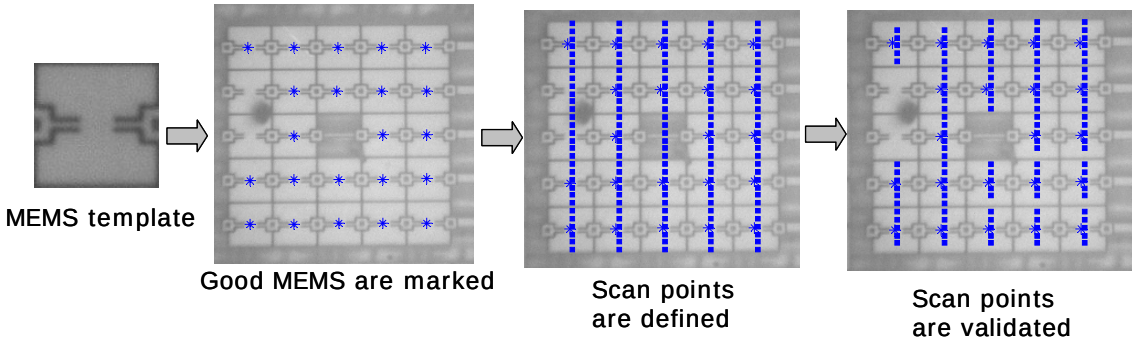


Figure 4: Procedure used to study the motion of MEMS using LDV.

Using a template of the device to be measured, the ‘good’ devices are marked. All missing devices or damaged devices (ex. particles on top) are excluded. Next the points were the LDV has to measure (scan points) are defined and validated. Points falling outside MEMS or on top of non-marked MEMS are disregarded. Figure 5 shows measurement results of the tilt of the micro-mirrors of the array shown in Fig. 4. From the measurement of the displacement of each mirror (Fig. 4, left) at different voltages, the tilt in function of voltage is obtained (Fig. 4, center) and the tilt at maximal voltage (Fig. 4, right). Devices that were indicated as being defective (see Fig. 4) are not measured.

Ex. 1016, Page 231

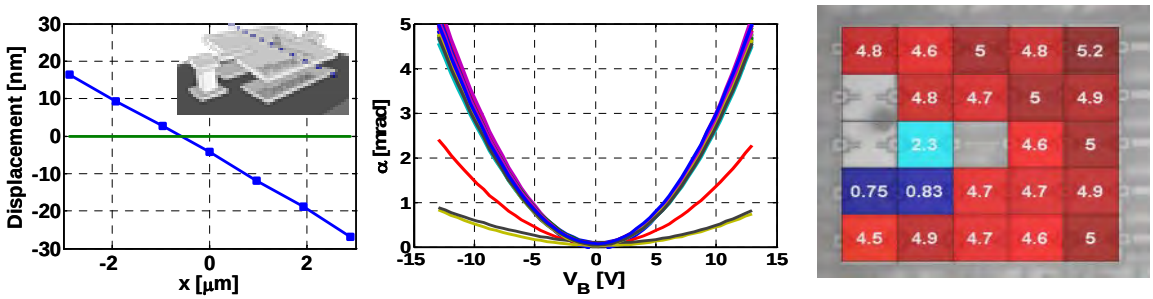


Figure 5: Left: Displacement of one micro-mirror at the positions of the probing laser beam. Center: Tilt in function of the applied voltage (center) for each tested MEMS. Right: Maximal tilt for the tested mirrors.

In addition to these automatic functionality measurements, the system was adapted to allow for long–term reliability experiments. The main challenge here was compensating for drift of the system which inevitably occurs during longer measurement time intervals. The out-of-plane drift (Z-drift) is corrected for using an autofocus system. XY-drift is corrected for by detecting the amount of drift and shifting the scan point grid by the same amount (5). Also for this procedure, a template is used. As a result, the scan sites effectively track the movement of the sample. This system allows monitoring very small devices during a very long period. For example, the resonance frequency of a 7.7 µm x 7.7 µm SiGe micro-mirror could be monitored in this way for almost $3x10^{12}$ cycles (> 400 hours). No degradation was seen, confirming the excellent reliability of these devices.

Acknowledgments

The authors thank the REMO team for help with measurements, and, especially, the whole SiGe micro-mirror processing team for developing and providing excellent devices for test and pushing the development of the techniques discussed in this paper.

References

1. M. Gromova, L. Haspeslagh, A. Verbist, B. Du Bois, R. Van Hoof, B. Eyckens, B. Sijmus, I. De Wolf, V. Simons, Ph. Muller, T. Lauwagie, M. Willegems, S. Locorotondo, W. Boullart, K. Baert and A. Witvrouw, *IEEE MEMS 2007*, 795 (2007)
2. L. Haspeslagh, J. De Coster, O. Varela Pedreira, I. De Wolf, B. DuBois, A. Verbist, R. Van Hoof, M. Willegems, S. Locorotondo, G. Bryce, J. Vaes, B. van Drieenhuizen and A. Witvrouw, in *Proc. IEDM 2008*
3. http://www.veeco.com/pdfs/appnotes/AN59-MEMS_Metro_04176_rf_247.pdf
4. O. Varela Pedreira, T. Lauwagie, J. De Coster, L. Haspeslagh, A. Witvrouw, I. De Wolf, in *Proc. 9th Int. Laser Metrology Symposium*, Singapore (2008)
5. J. De Coster, L. Haspeslagh, A. Witvrouw, I. De Wolf, in *Proc. 9th Int. Laser Metrology Symposium*, Singapore (2008)
6. K. Briechle and U.D. Hanebeck, in *Proc. SPIE, AeroSense Symposium*, Orlando, Florida, 1995, p. 4387 (1995)
7. I. De Wolf, J. De Coster, O. Valera Pedreira, L. Haspeslagh, A. Witvrouw, In *Proc. IEEE Sensors*, p. 73 (1008).
8. http://www.polytec.com/ger/default.asp

ECS Transactions, 18 (1) 205-211 (2009)
10.1149/1.3096451 © The Electrochemical Society

Reliability of Copper Interconnects: Stress-Induced Voids

Jeffrey P. Gambino, Tom C. Lee, Fen Chen, Timothy D. Sullivan,

IBM Microelectronics, Essex Junction, VT, 05494, USA

Stress-induced voids can form in Cu interconnects, due to either thermal expansion mismatch between the metal and the dielectric or due to confined grain growth in the Cu. The fail rate due to stress-induced voids increases as device dimensions decrease, because the critical void size to cause a fail decreases. Good process control is required for trench and via profiles, barrier and seed layer coverage, Cu fill, and cap layer adhesion, to prevent fails from stress-induced voids.

Introduction

Copper interconnects have gained wide acceptance in the microelectronics industry due to improved resistivity and reliability compared to Al interconnects (1). Initially, Cu interconnects were only used for high performance logic circuits. However, Cu interconnects are now used in a wide variety of integrated circuits, including dynamic random access memories (DRAM) (2) and CMOS image sensors (3). Ensuring the reliability of Cu interconnects becomes more challenging as device dimensions shrink. In this report, we will focus on stress-induced voiding in Cu interconnects. The first section describes the mechanism for the formation of stress-induced voids in Cu interconnects, while the second section discusses the effect of processing on formation of stress-induced voids.

Mechanisms for Stress-Induced Void Formation

Voids can form in passivated Cu interconnects during annealing at moderate temperatures (~ 200 to 250°C), due to tensile stress in the metal. There are two mechanisms for the stress in the metal; thermal stress, due to thermal expansion mismatch between the metal and the insulator, and growth stress, due to grain growth in the metal (4-6). Voids will form if the tensile stress is above the critical stress. If the voids grow large enough (i.e., spanning the width of a line or a via), the resistance will increase and the circuit will fail (Fig. 1). The lifetime for stress-induced voids is monitored using simple via chain structures, where the resistance is measured as a function of time at the stress temperature (Fig. 2).

Thermal stress in Cu wires is associated with high temperature processes (typically cap and interlevel dielectric deposition) after Cu CMP (4,6,7). After Cu CMP, the thermal stress in the wires is relatively low, because the metal processing occurs at low temperatures (typically < 300°C) and the top surface of the Cu is unconstrained. However, the deposition of the capping layers (SiN or SiCN) and interlevel dielectrics occurs at a relatively high temperature (> 300°C). Copper in the trenches expands due to heating at the deposition temperature, then contracts during cooling to room temperature. But the Cu is constrained by the dielectric capping layer, so it cannot shrink to the original dimensions (Fig. 3). As a result, there is a tensile stress in the Cu wires after deposition of the capping layer. Assuming there is no deformation of the dielectric during cooling, the stress in the z-direction, σ_z, in the Cu wire after capping layer deposition is given by the following:,

Ex. 1016, Page 233

$$\sigma_z = E_m \Delta\alpha \Delta T \qquad [1]$$

where E_m is the bulk modulus of the metal (120 GPa for Cu), $\Delta\alpha$ is the difference in thermal expansion mismatch between the metal and the surrounding dielectric layer (α_{Cu} = 17 ppm/^{o}K, α_{SiO_2} = 3 ppm/^{o}K), and ΔT is the difference in temperature between the dielectric deposition temperature and the temperature of interest (i.e. for example, the operating temperature of the device) (4). Hence, the stress in Cu will be higher as the temperature of the capping layer deposition increases.

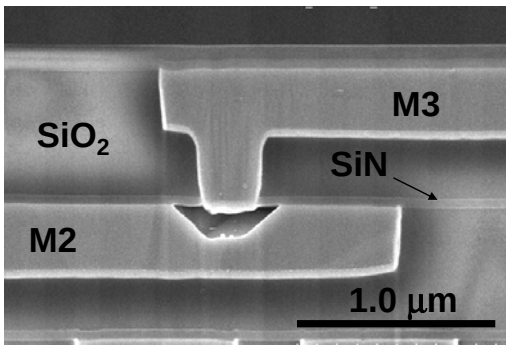


Fig. 1. SEM cross-section of failing via chain after stress migration test at 225°C for 1000 hr. Sample with 400°C post-metal anneal. M2 width = 20 µm (30).

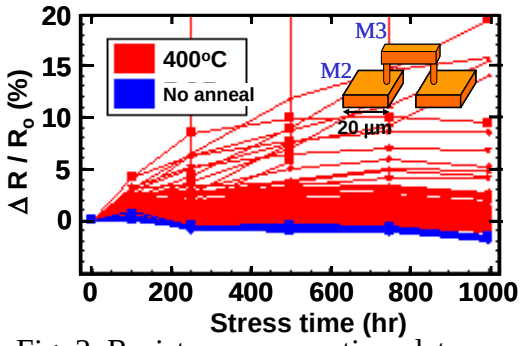


Fig. 2. Resistance versus time data during stress migration test for M3-V2-M2 via chain, for samples with or without post-metal anneal. M2 width = 20 µm, V2 diameter = 0.26 µm (30).

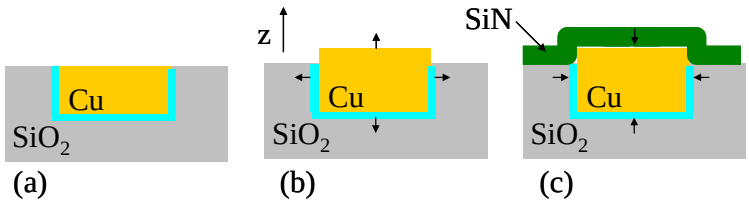


Fig. 3. Tensile stress in Cu due to cap deposition. (a) After Cu CMP (25°C); (b) During SiN deposition (~ 400°C); (c) After cooling back to room temperature.

Another contribution to tensile stress in Cu is confined grain growth (5). The copper plating process occurs at low temperatures (< 100°C), so that as-deposited Cu has a small grain size. Subsequent dielectric depositions occur at much higher temperatures (up to 400°C). Hence, grain growth will occur in the Cu. Grain growth eliminates excess free space in the Cu, resulting in excess vacancies and shrinkage of the metal lines (8). If grain growth occurs before cap layer deposition, the vacancies can diffuse to the surface and are annihilated. However, if grain growth occurs after cap deposition, the excess vacancies will be trapped in the Cu (Fig. 4). The shrinkage of the metal lines results in an additional component of tensile stress in the metal. To minimize void formation associated with confined grain growth, it is common practice in the industry to anneal the Cu after plating (5,9).

The tensile stress in the metal can be relieved by the formation of voids in the wire. If the voids grow large enough (i.e., causing a large enough resistance shift), the circuit will fail. In order for a void to form, the strain energy release (associated with void formation) must exceed the change in interface energy (i.e. assuming the void forms at an interface) (10-12). There exists a critical stress for void nucleation, above which a void will be thermodynamically stable, and growth is limited only be kinetics. The critical stress depends on the adhesion between Cu and the various barrier layers, and on the modulus of the dielectric. The critical stress is lower for Cu in a SiCOH

dielectric (E = 6 GPa) compared to Cu in F-doped SiO_2 (E = 71 GPa)(12,13). Hence, stress-induced voiding is more likely for Cu in SiCOH than for Cu in F-doped SiO_2 (even though the tensile stress is lower for Cu in SiCOH (12). The nucleation barrier for void formation is also reduced if there are pre-existing defects in the structure, such as undercut from the via process (9) or poor fill during the plating process (14).

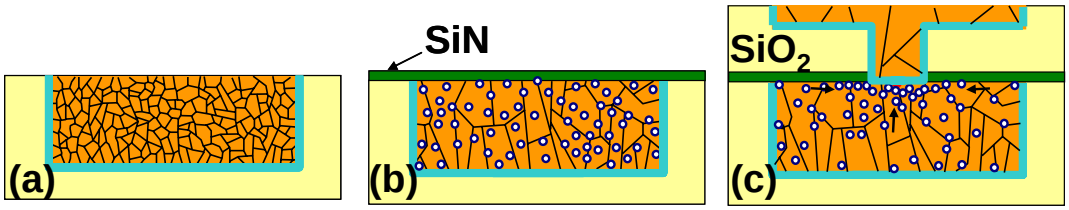


Fig. 4. Void formation in Cu due to confined grain growth. (a) Cu deposition and CMP, no anneal; (b) After cap deposition; metal is saturated with vacancies due to grain growth; (c) After via formation; void nucleates under via due to high tensile stress (5).

For stress-induced voids that originate from thermal expansion mismatch (i.e. during cap layer deposition), the maximum rate of void growth in Cu occurs below the "stress-free" temperature (5). The stress-free temperature is related to the deposition temperature of the cap layer and subsequent processes (such as interlevel dielectric deposition). At temperatures close to the stress-free temperature, the tensile stress, σ, in the metal is low, so the void growth rate is low. At low temperatures (< 150°C), the Cu diffusivity, D, is low, so the void growth rate is low (Fig.5). Significant void growth only occurs at intermediate temperatures (~200°C to 250°C), where there is both significant tensile stress in the metal and significant diffusivity of Cu. The rate of void growth, R, is given by the following (5):

$$R = C\,(T_o - T)^N \exp(-Q/kT) \qquad [2]$$

where T_o is the stress-free temperature, N is the creep exponent, Q is the activation energy, k is the Boltzman constant, and C is a proportionality constant. Measured values of the activation energy are typically ~ 0.75 eV. The activation energies for bulk Cu diffusion, grain boundary diffusion, and interface diffusion are 2.2 eV, 0.8 to 1.0 eV, and 0.8 to 1.1 eV, respectively (15). This suggests that the dominant diffusion mechanisms for stress-induced voiding are grain boundary and/or interfacial diffusion (5,16).

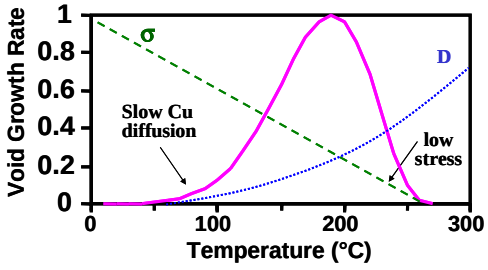


Fig. 5. Void growth rate as a function of temperature (5).

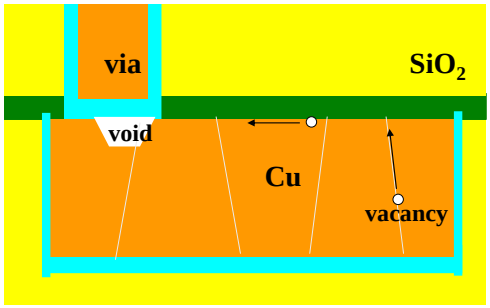


Fig. 6. Schematic of vacancy diffusion along fast diffusion paths.

Stress-induced voids in Cu are typically observed at grain boundaries (17), under vias (5), or inside vias (18). Voids form preferentially at grain boundaries and interfaces (i.e. often at the intersection of a grain boundary and the capping layer),

because these are fast diffusion paths compared to bulk diffusion (Fig. 6) (11). Void formation under vias is favorable due to the stress gradient in the underlying Cu and due to the presence of the via-metal interface. Modeling results show that for a via landing on a metal line, there is a high tensile stress in the metal at the edge of the via (5,12). A void will nucleate if the tensile stress exceeds the critical stress. If there is weak adhesion between the barrier metal and the underlying Cu at the bottom of the via, the critical stress will be reduced, making void nucleation even more favorable. Once a void nucleates, the stress field surrounding the void becomes less tensile, and the resulting stress gradient favors vacancy diffusion towards the void, resulting in further growth (19). Voids form in vias because the tensile stress in vias and narrow lines is lower than the tensile stress in wide lines. Vacancies will diffuse to the regions of lower tensile stress in the via. If there is a defect in the via (due to poor seed layer coverage, for example), then void nucleation will be further enhanced in the vias.

The line and via dimensions have a large effect on the failure rate due to stress-induced voids. The failure rate for stress-induced voids in Cu increases with increasing line width (Fig. 7), opposite to what is observed with Al (5,6). There are two reasons for this line-width dependence. For Cu lines, the hydrostatic stress increases with line width (due to higher growth stress) (6), so there is a larger driving force for void formation for wide lines compared to narrow lines. In addition, for wide lines, the kinetics of void formation are enhanced because a greater number of vacancies are available within a diffusion length of the via (Fig. 8)(5). In order to cause a fail, the via must be large enough to span the bottom of the via. The "active diffusion volume" is larger for wide lines than for narrow lines, so it takes less time to form a "killer void". The fail rate due to stress-induced voids increases as via size decreases (20) (Fig. 7), for similar reasons as for electromigration; The void size required to cause a fail decreases as the via size. Because stress-induced voids are mainly a problem for wide lines (rather than narrow lines), a simple design solution is to use redundant vias (21-23).

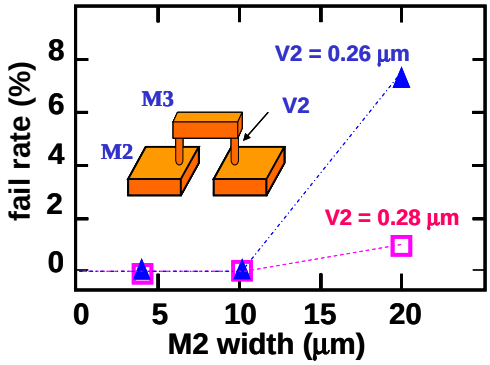


Fig. 7. Fail rate for stress-induced voids versus M2 line width at 225°C, 1000 h stress. Samples with 400°C post-metal anneal.

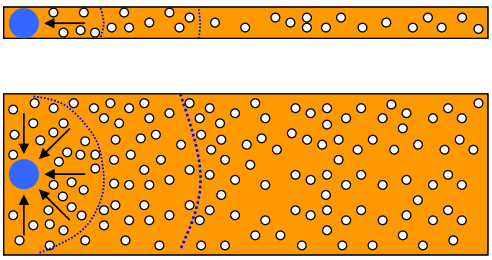


Fig. 8. Top-down schematic of vacancy diffusion towards via for narrow line versus wide line. For a given time, more vacancies are collected under the via in the wide line (i.e. more vacancies are within a diffusion length of the via).

The Effect of Processing on Stress-Induced Voids

Copper interconnects and vias are formed using a dual damascene process (1,24). The first step is deposition of the interlevel dielectric, followed by patterning of vias and trenches, that are etched into the dielectric. Next, the structure is cleaned, a barrier metal, such as TaN/Ta, and a Cu seed layer are deposited by ionized physical vapor deposition (I-PVD). Finally, the vias and trenches are filled using electroplated Cu, the excess metal is removed by chemical mechanical polishing (CMP), and a dielectric

capping layer is deposited. These steps are repeated for each metal level. After the last metal layer is fabricated, thick dielectric passivation layers are deposited and vias are opened to the bond pads. The chips are typically packaged using either a wirebond process or a flip-chip process.

To minimize the fail rate from stress-induced voids, it is important to have good metal barrier coverage on the bottoms and sidewalls of trenches and vias, to prevent Cu from diffusing into the dielectric. Having adequate barrier coverage is more difficult as dimensions shrink, due to the requirement to thin down the barrier layer (from 9 nm at the 65nm node to 3 nm at the 22 nm node) (25), to ensure low wire resistance. An additional challenge is ensuring adequate metal coverage on the sidewalls of porous low-k materials. Pore sealing prior to metallization is required to ensure good metal coverage (26). Because the barrier layers are deposited by physical vapor deposition , good control of the via and trench profiles is also critical in order to achieve adequate metal coverage (27).

The copper fill is also important to prevent failure due to stress-induced voids. If there are pre-existing voids because of incomplete Cu fill, the devices will fail after very short stress times. Achieving void-free Cu fill is more challenging as device dimensions scale down, because the Cu seed layer thickness must be reduced (to prevent pinch-off at the tops of trenches and vias) and because the resulting aspect ratio for Cu plating is increased. Improvements to the PVD tools are required as device dimensions scale down, to ensure good uniformity of the Cu seed layer across the wafers (28). In addition, the Cu plating process must be optimized, for example, by optimizing the additives in the plating bath (29).

Annealing also affects the formation of stress-induced voids. An anneal is required after Cu plating to maximize grain size prior to cap layer deposition (i.e. to prevent confined grain growth). But after cap layer deposition, the maximum anneal temperature must be limited. High temperature anneals after cap deposition can lead to high rates of stress-induced void formation (Fig. 2), due to either confined grain growth or due to increased stress in the Cu (17, 30).

Because the kinetics of void formation are controlled by interface and grain-boundary diffusion (31), the rate of void growth can be reduced by using metal capping layers (32) (Fig. 9) or by alloying the Cu (33,34) (Fig. 10), similar to methods used to improve electromigration lifetime. As with electromigration, it is important to form a strongly adhering interface between the Cu and the capping layer, in order to minimize vacancy diffusion along this fast-diffusion path.

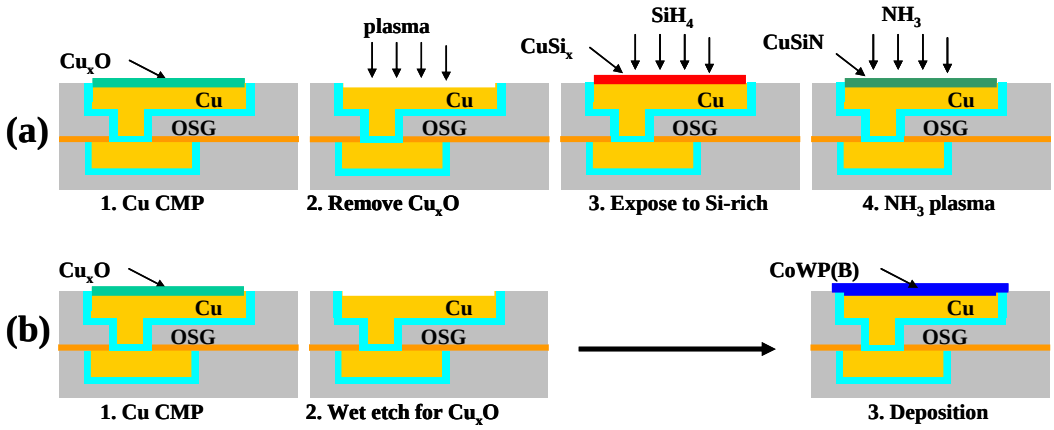


Fig. 9. Process flow for self-aligned cap layer on Cu using (a) SiH_4 exposure or (b) electroless metal deposition.

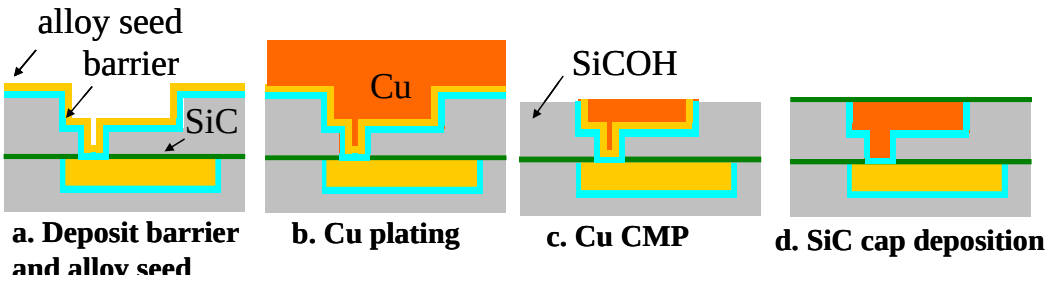


Fig. 10. Process flow for alloying Cu using alloy seed layer.

Conclusion

Stress-induced voids can form in Cu interconnects, due to either thermal expansion mismatch between the metal and the dielectric or due to confined grain growth in the Cu. The fail rate due to stress-induced voids increases as device dimensions decrease, because the critical void size to cause a fail decreases. Good process control is required for trench and via profiles, barrier and seed layer coverage, Cu fill, and cap layer adhesion, to prevent fails from stress-induced voids.

Acknowledgements

The authors acknowledgement Carl Farris and the staff of IBM Burlington manufacturing facility for help with processing the wafers.

References

1. D. Edelstein, J. Heidenreich, R. Goldblatt, W. Cote, C. Uzoh, N. Lustig, P. Roper, T. McDevitt, W. Motsiff, A. Simon, J. Dukovic, R. Wachnik, H. Rathore, R. Schulz, L. Su, S. Luce, J. Slattery, *IEDM Proc.*, 1997, p. 773.
2. H.B. Lee, J.W. Hong, G.J. Seong, J.M. Lee, H. Park, J.M. Baek, K.I. Choi, B.L. Park, J.Y. Bae, G.H. Choi, S.T. Kim, U.I. Chung, J.T. Moon, J.H. Oh, J.H. Son, J.H. Jung, S. Hah, S.Y. Lee, *IITC Proc.*, 2007, p. 64.
3. J. Gambino, J. Adkisson, T. Hoague, M. Jaffe, R. Leidy, R.J. Rassel, J. Kyan, D. McGrath, D. Sackett, Proc. *AMC 2005*, MRS, 2006, p. 151.
4. T.D. Sullivan, *Ann. Rev. Mat. Sci.*, 26, 1996, p. 333.
5. E.T. Ogawa, J.W. McPherson, J.A. Rosai, K. J. Dickerson, T.-C. Chiu, L.Y. Tsung, M.K. Jain, T.D. Bonifield, J.C. Ondrusek, *IRPS Proc.*, 2002, p. 312.
6. J.-M. Paik, J.-K. Jung, Y.-C. Joo, *IRPS Proc.*, 2005, p. 195.
7. Z.-S. Choi, R. Monig, C.V. Thompson, *J. Mater. Res.*, **23**, 2008, p. 383.
8. Y.-C. Joo, J.-M. Paik, J.-K. Jung, *MRS. Proc.*, **vol. 863**, 2005, p. 241.
9. A.H Fischer, A. von Glasow, S. Penka, F. Ungar, *IITC Proc.*, 2003, p. 253.
10. P.A. Flinn, S. Lee, J. Doan, T.N. Marieb, J.C. Bravman, M. Madden, *in Stress Induced Phenomena in Metallization*, 1998, p. 250.
11. R.J. Gleixner, B.M. Clemens, W.D. Nix, *J. Mater. Res.*, **12**, 1997, p. 2081.
12. C.J. Zhai, H.W. Yao, P.R. Besser, A. Marathe, R.C. Blish II, D. Erb, C. Hau-Riege, Sidharth, K.O. Taylor, *IRPS Proc.*, 2004, p. 234.
13. C.S. Hau-Riege, S.P. Hau-Riege, A.P. Marathe, J. Appl. Phys., **96**, 2004, p. 5792.
14. K. Arita, N. Ito, N. Hosoi, H. Miyamoto, *IEEE Semicond. Manu. Symp.*, 2001, p. 155.
15. C.-K. Hu, L. Gignac, R. Rosenberg, *Microelec. Rel.*, **46**, 2006, p. 213.
16. D. Gan, B. Li, P.S. Ho, *MRS Proc.*, vol. **863**, 2005, p. 259.

17. T.M. Shaw, L. Gignac, X.-H. Liu, R.R. Rosenberg, E. Levine, P. McLaughlin, P.-C. Wang, S. Greco, G. Biery, *in Stress-Induced Phenomena in Metallization*, 2002, p. 177.
18. K.Y.Y. Doong, R.C.J. Wang, S.C. Lin, L.J. Hung, S.Y. Lee, C.C. Chiu, D. Su, K. Wu, K.L. Young, Y.K. Peng, *IRPS Proc*., 2003, p. 156.
19. S. Orain, A. Fuchsmann, V. Fiori, X. Federspiel, *Proc. EuroSime*, 2006, p. 1.
20. T. Oshima, K. Hinode, H. Yamaguchi, H. Aoki, K. Torii, T. Saito, K. Ishikawa, J. Noguchi, M. Fukui, T. Nakamura, S. Uno, K. Tsugane, J. Murata, K. Kikushima, H. Sekisaka, E. Murakami, K. Okuyama, T. Iwasaki, *IEDM Proc*., 2002, p. 757.
21. K. McCullen, *ISQED Proc*., 2007, p. 821.
22. K. Yoshida, T. Fujimaki, K. Miyamoto, T. Honma, H. Kaneko, H. Nakazawa, M. Morita, *IEDM Proc*., 2002, p. 753.
23. G.D.R. Hall, D.D.J. Allman, H.D. Bhatt, *IRPS Proc*., 2008, p. 392.
24. J. Gambino, A. Stamper, T. McDevitt, V. McGahay, S. Luce, T. Pricer, B. Porth, C. Senowitz, R. Kontra, M. Gibson, H. Wildman, A. Piper, C. Benson, T. Standaert, P. Biolsi, E. Cooney, E. Webster, R. Wistrom, A. Winslow, E. White, *IPFA Proc*., 2002, p. 111.
25. M. Gallitre, L.G. Gosset, A. Farcy, B. Blampey, R. Gras, C. Bermond, B. Flechet, J. Torres, *IITC Proc*., 2007, p. 132.
26. S. Arakawa, I. Mizuno, Y. Ohoka, K. Nagahata, K. Tabuchi, R. Kanamura, S. Kadomura, *IITC Proc*., 2006, p. 210.
27. F.Chen, B. Li, T. Lee, C. Christiansen, J. Gill, M. Angyal, M. Shinosky, C. Burke, W. Hasting, R. Austin, T. Sullivan, D. Badami, J. Aitken, *IPFA Proc*., 2006, p. 97.
28. N. Kumar, K. Moraes, M. Narasimhan, P. Gopalraja, *Semicond. Int*., May 2008, p. 33.
29. Y.-C. Huang, X. Lin, B. Zheng, C.S. Ngai, V. Paneccasio, J. Behnke, C. Witt, J. Dukovic, A. Rosenfeld, *AMC Proc*., 2005, p. 507.
30. J. Gambino, T.C. Lee, D. Meatyard, S. Mongeon, B. Li, F. Chen, *ISTC Proc*., 2008, p. 13.
31. D. Gan, P.S. Ho, Y. Pang, R. Huang, J. Leu, J. Maiz, T. Scherban, *J. Mater. Res*., **21**, 2006, p. 1512.
32. T. Ishigami, T. Kurokawa, Y. Kakuhara, B. Withers, J. Jacobs, A. Kolics, I. Ivanov, M. Sekine, K. Ueno, *IITC Proc*., 2004, p. 75.
33. K. Maekawa, K. Mori, K. Kobayashi, N. Kumar, S. Chu, S. Chen, G. Lai, D. Diehl, M. Yoneda, *AMC Proc*., 2004, p. 221.
34. S. Yokogawa, H. Tsuchiya, *J. Appl. Phys*., **101**, 2007, p. 013513

ECS Transactions, 18 (1) 213-218 (2009)
10.1149/1.3096452 © The Electrochemical Society

Si-O bond-breakage energetics under consideration of the whole crystal

S.E. Tyaginov[a], V. Sverdlov[b], W. Gös[a], Ph. Schwaha[b],
R. Heinzl[b], F. Stimpfl[b], T. Grasser[a]

[(a)]Christian Doppler Laboratory for TCAD at the [(b)]Institute for Microelectronics,
TU Wien, Gußhausstraße 25-29, A-1040 Vienna, Austria

We extend the McPherson model for Si-O bond-breakage energetics, which considers only a single SiO_4 tetrahedron, in a way to capture the effect of the whole crystal. No ledge in the direction assumed in the McPherson model was revealed. Instead, potential profiles feature a saddle point in a different direction. The activation energy for bond-breakage is found to be rather high (~ 6 eV). This suggests that the interaction with the electric field alone is not sufficient for bond rupture. The breakage rate calculated for a bond weakened by hole capture is comparable to that of the McPherson model. We conclude that only the common action of an electric field and other factors (structural disorder and/or energy delivered by particles, e.g. hot carriers) can result in bond-breakage.

1. Introduction

Two main concepts of the breakdown of SiO_2 film exist in the literature: the first one relates the dielectric degradation with energy delivered by particles and the second one treats the interaction with the electric field as the triggering mechanism for oxide damage. One may distinguish two main approaches supporting the first concept – namely the Anode Hole Injection model (1,2) and the Anode Hydrogen Release model (3,4) – while the Electro-Chemical model (5,6) treats the Si-O bond-breakage in terms of a chemical reaction where the activation energy is lowered by the field. At the same time, the rupture of Si-O bonds has been suggested to be an essential component of the Hot Carrier induced degradation (7,8) and of the Time Dependent Dielectric Breakdown (5,6) and thus the model of Si-O bond breakage is the one of the crucial challenges in the area of SiO_2 film reliability. The McPherson model (5,6) treats the rupture of Si-O bonds as a transition of the Si ion from the 4-fold position to the 3-fold coordination. This model, however, considers only a single SiO_4 tetrahedron. It is obvious that the contribution of the whole surrounding network substantially changes the situation. For instance, neighboring Si ions are situated in a distance of ~ 3.2 Å from the center of the tetrahedron and carry the charge twice as large as that of the O ions. Thus the contribution to the energy of each Si ion is comparable to that provided by the O ions. In this work the McPherson model will be extended in a way to capture the effect of the whole lattice.

2. Generalization of Mie-Grüneisen Potential

Following McPherson, we employ the Mie-Grüneisen potential (MGP) (5,6) to describe pair-wise interatomic interactions in SiO_2. However, we take into account not only Si-O, but also Si-Si and O-O interactions and use the MGP in a generalized form:

$$\Phi_{ij}(r_{ij}) = Q_i^{(9)} Q_i^{(9)} r_{ij}^{-9} + Q_i^{(2)} Q_i^{(2)} r_{ij}^{-2} + Q_i^{(1)} Q_i^{(1)} r_{ij}^{-1}, \qquad [1]$$

Ex. 1016, Page 241

where the indices *i*,*j* indicate the type of the atom, r_{ij} is the interatomic distance and $Q_i^{(m)}$ are "effective charges" of ions corresponding to the term of the *m*-th power. We have 6 constants to be determined (2 types of ions with each characterized by 3 parameters).

The number of independent constants is further reduced by the requirement of finiteness of the electrostatic energy (for each *m*):

$$U^{(m)} = \frac{1}{2} \sum_{n_1,n_2,n_3} \sum_i \sum_j \frac{Q_i^{(m)} Q_j^{(m)}}{|\vec{r}_i - \vec{r}_j + \vec{n}|^m}, \qquad [2]$$

where *i*,*j* enumerate ions in a primitive cell, $\boldsymbol{r}_i$, $\boldsymbol{r}_j$ are the positions of the atoms and $\boldsymbol{n} = n_2\boldsymbol{t}_2 + n_2\boldsymbol{t}_2 + n_3\boldsymbol{t}_3$ with $\{\boldsymbol{t}_1,\boldsymbol{t}_2,\boldsymbol{t}_3\}$being basis translation vectors. These series converge only when the unit cell is neutral, i.e. presuming that it contains 3 Si and 6 O atoms, one writes $Q_{Si}^{(m)} / Q_O^{(m)} = -2$ for each *m* and as a consequence only 3 independent constants remain.

The McPherson model uses 3 constants, which are found by the condition that the potential energy has a minimum at the Si equilibrium position (EP) (zero force on the Si ion and the energetic position of the minimum is equal to the bond strength) and the bond polarity must represent the conventional value for the Si-O bond. The constant set used in the original model is not applicable in the extended version because it results in a wrong value of the cohesion energy. Also the potential acting on the O ion reveals an energy maximum instead of a minimum at the EP. To determine $Q_i^{(m)}$ a calibration scheme reproducing some material characteristics (cohesion energy and elastic constants) has to be employed. Another suitable way is to calibrate the MGP results to the data obtained by the Density Functional Theory (DFT) combined with the Molecular Dynamics (MD). However, a more straightforward way is to employ TTAM and BKS interatomic potentials which have been designed just to reproduce results of DFT/MD (9-12).

3. TTAM and BKS potentials

The general expression for TTAM and BKS potentials contains 3 terms: the Coulombic, the covalent, and the short-range repulsive term precluding the collapse of atoms:

$$\Phi_{ij}(r_{ij}) = \frac{kQ_iQ_j}{r_{ij}} + \alpha_{ij} e^{-\beta_{ij} r_{ij}} - \gamma_{ij} r_{ij}^{-6}. \qquad [3]$$

The constant *k* converts $e^2Å^{-1}$ to eV. As in [1], indices *i*,*j* enumerate the types of the atoms. We employ only the versions of potentials based on commonly-accepted values of Q_{Si}/Q_O = +2.4/-1.2 (in electron charge) as effective charges in the Coulombic term. Diverse forms of TTAM/BKS differ in values of constants α_{ij}, β_{ij}, γ_{ij} listed in the Table I (9-12).

The potential acting on the Si ion from the rest of the crystal vs. its displacement from the EP perpendicular to the O_3 plane is shown in Fig. 1a and toward the middle of the O-O segment is depicted in Fig 1b. In the direction declared in the McPherson model no saddle point is observed. Instead, the secondary minimum/ledge is revealed for the

Ex. 1016, Page 242

second direction for all types of potentials. The activation energy needed for the Si atom to transfer from the 4-fold coordination to the 3-fold position is rather high, i.e. ~ 6 eV vs. ~ 2.3 eV found in (5,6). This circumstance is related to the effect of the surrounding network, primarily, to the presence of 4 neighboring Si ions providing a strong positive contribution to the potential. Although energy positions of the primary minimum and the ledge are different for various potentials, the activation energies for bond-breakage are approximately the same. We use the classical version of the TTAM potential further on.

TABLE I. Values of constants used in various versions of TTAM/BKS potentials (9-12)

Version of the potential	α_{ij}	β_{ij}	γ_{ij}
Si-Si TTAM	$8.7235 \cdot 10^8$	15.2207	23.30
Si-Si BKS	0	0	0
Si-Si Z1-TTAM	0	0	0
Si-Si FB-TTAM	$7.950 \cdot 10^4$	4.975	446.780
Si-O TTAM	10721.5	4.7959	70.7343
Si-O BKS	18003.8	4.8738	133.538
Si-O Z1-TTAM	7149.0	4.7864	27.661
Si-O FB-TTAM	10450.0	4.8077	63.047
O-O TTAM	1756.90	2.8464	214.736
O-O BKS	1388.773	2.760	175.00
O-O FB-TTAM	1428.0	2.7933	41.374
O-O Z1-TTAM	1359.0	2.8086	215.829

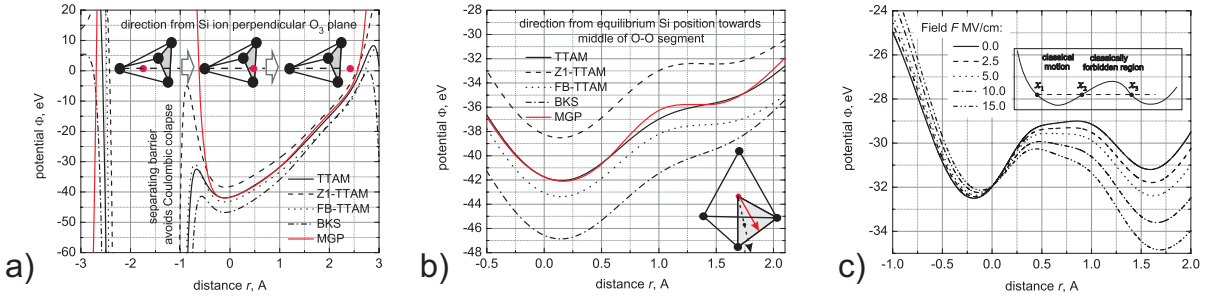


Figure 1. The potential acting on the Si ion vs. its displacement from the EP: (a) perpendicular to the O_3 plane (see inset); (b) toward the middle of the O-O segment (inset: red arrow, black arrow – McPherson direction); (c) transformation of the potential with the electric field (bond is weakened by HC).

4. Calibration of Mie-Grüneisen Potential

For both types of ions the potential acting on them should reveal a minimum at their EP. We employ the energy positions of the corresponding minima for Si and O ions as well as the cohesion energy of α-SiO_2 (the values published in the literature span the range from 18 to 19.1 eV/formula unit (13-15); 19.1 eV/formula unit, Ref. (15) is used in this work) to find 3 constants in the MGP. It is more convenient to find the products $Q_{Si}^{(m)}Q_O^{(m)}$ ($Q_O^{(m)}Q_O^{(m)} = -1/2\, Q_{Si}^{(m)}Q_O^{(m)}$, $Q_{Si}^{(m)}Q_{Si}^{(m)} = -2\, Q_{Si}^{(m)}Q_O^{(m)}$) rather then the $Q_{Si}^{(m)}, Q_O^{(m)}$ separately. Table II compares the values found in this work with those from the McPherson model. Note that we have rewritten the dimensionless constants used in (5) in terms of $Q_{Si}^{(m)}Q_O^{(m)}$.

The potential for the Si ion calculated with properly-parameterized MGP is shown in Fig. 1a-b for the direction assumed in the McPherson model and that revealed in the present extended version, respectively. One can see that also with MGP employed no ledges are revealed in the first case while in the direction from the Si EP towards the middle of the O-O segment the secondary minimum is observed at the same positions as with

TTAM/BKS. We intentionally did not include the secondary minimum into the calibration procedure to avoid its artificial appearance predefined by the calibration.

TABLE II. Comparison of the MGP constants employed in the McPherson model and those used in the present model.

Power of term *m*	McPherson Model	Present model
9	117.964 eV·Å^9	99.480 eV·Å^9
2	-7.392 eV·Å^2	-3.148 eV·Å^2
1	-6.523 eV·Å	-41.308 eV·Å

5. Effect of hole capture and electric field

As it was reported in the original McPherson model, the effective dipole moment lies in the range of 7…13 *e*Å and thus the dipole – electric field interaction could change the potential energy by ~ eV. Keeping in mind the huge activation energy of ~ 6 eV for bond-breakage obtained in our model, we conclude that the electric field ***F*** only is not sufficient for the rupture of Si-O bonds and the contribution of other factors is essential for effective bond-breakage. One of the possible candidates may be a bond weakening by hole capture (HC) and, thus, we consider the transformation of the Si binding potential with the electric field for a bond distorted by HC.

We assume a captured hole is situated in the middle of the Si-O bond and its contribution into the potential ($\boldsymbol{r}_{Si}/\boldsymbol{r}_O$ are the position of the Si/O ion, $\varepsilon_0\varepsilon_I = 3.9$ is the permittivitiy) is:

$$\Phi_{hole} = \frac{+|e|\cdot Q_{Si}}{4\pi\varepsilon_o\varepsilon_I|\vec{r}_{Si} - 1/2(\vec{r}_{Si} - \vec{r}_O)|} = \frac{+|e|\cdot Q_{Si}}{8\pi\varepsilon_I\varepsilon_0|\vec{r}_{Si} - \vec{r}_O|}, \qquad [4]$$

With Si being in the EP the dipole moment $\boldsymbol{p} = \mathbf{0}$ but it develops as Si is shifted. Since the Si/O ions have a charge of $+4|e|f_i^*/-2|e|f_i^*$ (f_i^* is the bond polarity; *i* enumerates 4 O atoms bonded to the Si) the contribution to the potential due to the interaction with the field is:

$$\Phi_{dip}(\vec{r}) = -\vec{p}\cdot\vec{F} = -2|e|\left[\sum_{i=1}^{4} f_i^*(\vec{r} - \vec{r}_i)\right]\left(\frac{2+\varepsilon_I}{3}\right)\vec{F}. \qquad [5]$$

Similarly to the McPherson model, we assume that the bond polarity changes if the Si ion is displaced from its EP as $f_i^* = \dfrac{f_o^*}{(1-f_o^*)(r_0/r) + f_o^*}$, where f_0^* is the bond polarity in equilibrium, $r_0 = 1.7$ Å is the bond length and *r* is the ion displacement.

The transformation of the potential profile $V(x)$ of the Si ion calculated for a bond weakened by HC with the applied field *F* is depicted in Fig. 1c. The degeneracy of the two minima occurs at $F_{cr} \sim 5$ MV/cm. Comparing this value with one reported in the McPherson model (15 MV/cm) one concludes that our critical field is more realistic, because 15 MV/cm is even higher than the dielectric strength of silica ~ 10 MV/cm (16).

6. Quantization effects and bond-breakage probability

Following McPherson, we treat the rupture of the Si-O bond as a transition of the Si ion from the 4-fold coordinated position to the 3-fold coordination outside the SiO_4 tetrahedron. We assume that two processes contribute to bond-breakage: thermal activation of the Si ion over the barrier separating the primary and the secondary minima

Ex. 1016, Page 244

and its tunneling through the barrier as in the McPherson model. The second mechanism will be treated quasi classically, i.e. within the WKB approximation.

In the quantum well corresponding to the primary minimum there is a system of energy levels; each level is characterized by its energy E_n, occupancy f_n, allez-retour time $\tau_{a\text{-}r,n}$, and the probability T_n to tunnel through the barrier are found from:

$$\int_{x_1}^{x_2} \sqrt{2m_{Si}\left(E_n - V(x)\right)}dx = (n+1/2)\pi\hbar\,, \qquad [6a]$$

$$f_n = \exp(-E_n / k_B T) / \sum_n \left[\exp(-E_n / k_B T)\right] \qquad [6b]$$

$$\tau_{a-r,n} = 2\sqrt{\frac{m_{Si}}{2}}\int_{x_1}^{x_2} \frac{dx}{\sqrt{E_n - V(x)}} \qquad [6c]$$

$$T_n = \exp\left(-\frac{2}{\hbar}\int_{x_2}^{x_3} \sqrt{2m_{Si}\left[V(x) - E_n\right]}dx\right), \qquad [6d]$$

m_{Si} is the mass of the Si ion, $\{x_1,x_2\}$ and $\{x_2,x_3\}$ are points limiting the classically allowed and prohibited regions of Si motion (Fig. 1c, inset). The tunneling rate from each level is expressed as $P_n = f_n T_n/\tau_{a\text{-}r,n}$. Note that the contribution to the bond-breakage process by thermionic emission is negligibly small with respect to tunneling, because the relatively high and narrow barrier is rather "suited" for tunneling than for thermal activation.

The level position as a function of its quantum number is presented in Fig. 2a; the number of levels decreases with F when the quantum well becomes shallower. Since tunneling occurs only from levels situated above the bottom of the 2nd minimum the offset of tunneling is observed in Fig 2b which shows P_n as a function of F. The trade-off between rapidly increasing with E_n tunneling transparency T_n and decreasing occupancy f_n reveals that the main contribution to the total rate P_n is due to 2-3 first levels. Fig 2c compares the total bond-breakage rate calculated in the present model for a bond weakened by HC and that borrowed from (6) (multiplied by an attempt rate $\nu \sim 10^{13}$ s).

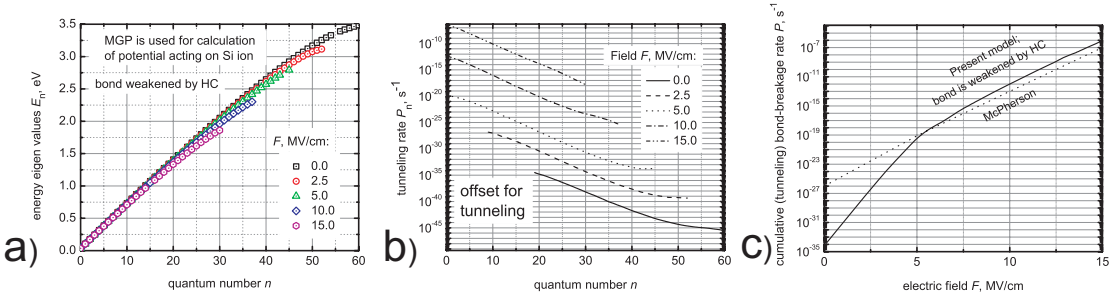


Figure 3. Bond-breakage energetics: (a) the system of energy levels; (b) tunnel rate vs. the level number for various F; (c) bond-breakage rates obtained with the McPherson model for a "virgin" bond vs. that obtained in our model for a bond weakened by HC.

Fig. 2c demonstrates similar rates obtained for 2 principally different cases: with and without bond weakening by HC. The McPherson model considers only a single tetrahedron. This suggests that under the consideration of the whole crystal the

interaction with an electric field could not provide a considerable bond-breakage rate and only its combined action with other factors such as structural disorder (HC, bond angle variations, bond stain) and/or energy delivered by particles (hot carriers, mobile hydrogen) results in bond rupture.

7. Conclusion

An extended version of the McPherson model which captures the effect of the whole crystal on Si-O bond-breakage energetics has been developed. Similar results were obtained with several types of pair-wise potentials. It was revealed that the surrounding network strongly stabilizes the SiO_4 tetragonal bonding configuration. No saddle point in the McPherson direction was observed. Instead a saddle point appears in the direction connecting the center of the SiO_4 tetrahedron and the middle of the O-O segment. The degeneracy of the primary and the secondary minima occurs at lower fields than in the McPherson model, i.e. at ~ 5 MV/cm. The barrier separating the two minima is too high for the thermal activation. Therefore, the interaction of the dipole moment with an electric field is not able to provide a significant rate for bond-breakage. Thus, the common action of a field and other factors is required for bond rupture. Among these factors are bond weakening by hole capture, bond angle deviations, bond strain, etc as well as energy deposited by particles, primary by hot carriers and mobile hydrogen. It is shown that the bond-breakage rate calculated within the McPherson model for a "virgin bond" and within the present model for a bond weakened by hole capture can be similar.

References

1. M. Alam, J. Bude, A, Ghetti, *38th Intern. Reliab. Phys. Symp.*, p. 21 (2000).
2. J. Bude, B. Weir, P. Silverman, *IEDM Tech. Dig.*, p. 179 (1998).
3. D.J. DiMaria, J.H. Stathis, *J. Appl. Phys.*, **89**, 9, p. 5015 (2001).
4. A. Ghetti, M. Alam, J. Bude, D. Monroe, E. Sangiorgi, H. Vaidya, *IEEE Trans. Electron Dev.*, **47**, **7**, p. 1341 (2000).
5. J.W. McPherson, *J. Appl. Phys.*, **99**, 8, pap. No. 083501, 13 pages (2006).
6. J.W. McPherson, *45th Inter. Reliab. Phys. Symp.*, p. 209 (2007).
7. S. Mahapatra, D. Saha, D. Varghese, P.B. Kumar, *IEEE Trans. Electron Dev.*, **53**, 7, p. 1583 (2006).
8. D. Saha, D. Varghese, S. Mahapatra, *IEEE Electron Dev. Lett.*, **27**, 7, p. 585 (2006).
9. S. Tsuneyuki, M. Tsukada, H. Aoki, Y. Matsui, *PRL.*, **61**, 7, p. 869 (1998).
10. B.W.H. van Beest, G.J. Kramer, R.A. van Santen, *PRL*. **64**, 16, p. 1955 (1998).
11. A.R. Al-Derzi, M.G. Gory, K. Runge, S.B. Trickey, *J. Phys. Chem*. A, **108**, 52, p. 11679 (2004).
12. W. Zhu, K. Runge, S.B. Trickey, *J. Comp.-Aid. Mater. Design*, **13**, 1-3, p. 75 (2006).
13. A. Tandia, G. Sarrabayrouse, A. Martinez, *Thin Solid Films*, **296**, 1-2, p. 122 (1997).
14. F. Jollet, C. Noguera, *Phys. Stat. Sol. (b)*, **179**, 2, p. 473 (1993).
15. Sh. Munetoh, T. Motooka, K. Moriguchi, A. Shintani, *Comput. Mat. Sci.*, **39**, 2, p. 334 (2007).
16. S.M. Sze, K.K. Ng, *Physics of Semiconductor Devices*, 3rd edition, Wiley and Sons, New-York, 2006.

Ex. 1016, Page 246

ECS Transactions, 18 (1) 219-223 (2009)
10.1149/1.3096453 © The Electrochemical Society

Drain Bias Effect on the Interface Traps of pMOSFETs under Negative Bias Temperature Stress

J.Y. Pan[a], J.Q. Yang[a], Y. Qiao[a], X.Y. Liu[a], R.Q. Han[a], *J.F. Kang[a],
C. C. Liao[b], H.M. Wu[b], Y.J. Wu[b]
[a] Institute of Microelectronics, Peking University, Beijing 100871, P. R. China
*Email: kangjf@pku.edu.cn
[b]Semiconductor Manufacturing International Corporation, 18 Zhangjiang Road, Shanghai

Keywords: Negative bias temperature instability (NBTI), charge-pumping technique (CP), interfacial traps, nitrided oxides.

Abstract

Negative bias temperature instability (NBTI) has been a major reliability issue for advanced CMOS technology. Significant drain-bias effect on NBTI has been demonstrated. In this paper, the correlation between drain-bias dependent NBTI and the interface trap (N_{it}) is studied. Charging pumping measurement was performed to monitor the N_{it} behavior and on-the-fly technique was used to measure the degradation and recovery behaviors of threshold voltage (ΔV_{th}) in the pMOSFETs. The results show that significant NBTI degradation under higher drain bias can be owed to the drain bias enhanced generation of the non-recovered interfacial traps associated with Si-H bond breaking effect.

1. Introduction

Negative bias temperature instability (NBTI) has been a serious reliability issue for advanced CMOS technology. It is widely accepted that threshold voltage shift is primarily due to Si/SiO_2 interface trap generation and hole trapping. Recently, significantly drain bias dependent NBTI behaviors were reported [1, 2, 3]. Therefore it is interesting to understand drain bias on NBTI degradation mechanism. In this paper, the correlation between drain-bias dependent NBTI and the interface trap (N_{it}) is addressed. Charge pumping (CP) technique [4] is used to measure the interface traps characterization of pMOSFETs under DBNBTI. We report and demonstrate that the interface state generation behavior is similar to that of ΔV_{th}. The drain bias dependent NBTI is due to Si-H bond breaking and mainly cause the non-recovered interfacial traps.

2. Experimenta Details

Poly-Si/SiON gated pMOSFETs with 0.5*um* channel length were applied by using 65*nm* technology. The corresponding EOT is 5.4*nm*. The oxynitride gate dielectrics were grown by in-situ steam generated oxidation followed by decoupled plasma nitridation and post-deposition annealing. NBTI stress is applied to the devices with gate bias stress ($V_{gstress}$) of -5.5*V* and various drain biases (ranging from 0.0 to -4.5*V*). By using standard on-the-fly (OTF) [5], I_{dlin} measurements were done to monitor the degradation of threshold voltage (ΔV_{th}). The generation of interface trap was measured by improved charge pumping methodology. The setup of CP technique is shown as follow: the

Ex. 1016, Page 247

source and drain are grounded; a pulse of fixed amplitude 2*V* was applied at the gate; the rising and falling rates of square wave were held at a constant value of 200*V/μs*; the frequency of gate pulses was fixed at 1*MHz*. The stress and measurement temperature is 300*K* and stress time is 500*sec*.

3. Results and Disscusion

A. the threshold voltage degradation

The drain voltage dependency of threshold voltage degradation for 0.50*um*-channel length pMOSFETs at gate stress voltage of -5.5*V* with various drain voltage are shown in Figure 1. The threshold voltage shift initially decreases and subsequently turns around to increase when $|V_{ds}|$ increases. The observed slopes of V_{th} shift versus stress time vary between 0.155-0.195 with drain bias increased, as shown in Figure 2. Since HCI has an exponent around 0.5, this indicated that even with high drain bias the degradation mechanism is governed by ≡Si-H bond breaking and H_2 diffusion away from Si-SiO_2 [6]. In the low drain bias region, as V_{ds} reduced E_{ox}, the NBTI degradation would be reduced. However, when the V_{ds} above a certain value as hot carriers arrive the saturation velocity, high energy holes equally lower the thermal barrier of tunneling to ≡Si-H bond in the surface, and reduced the activation energy of ≡Si-H bond dissociation [1], so larger magnitude of N_{it} generated.

B. the Nit generation

The CP technique is an efficient tool to extract the interface traps [4]. Figure 3 shows the 1*MHz* variable base-level charge pumping results of devices under 500*s* DBNBTI stress. The interface trap generation of device stressed at high gate bias and various drain bias voltage is plotted in Figure 4. It demonstrates that N_{it} generation curves have a similar turnaround degradation behavior as ΔV_{th}, which also imply that threshold voltage degradation is mainly due to Si-SiO_2 interface trap generation with Si-H bond breaking.

Compared with $\Delta V_{th}(V_{ds}=-4.5V)/\Delta V_{th}(V_{ds}=0.0V)$, the value of $\Delta N_{it}(V_{ds}=-4.5V)/\Delta N_{it}(V_{ds}=0.0V)$ is much larger and the turnaround voltage for ΔN_{it} curve almost 0.5V smaller. These can be explained by the recovery ratio. Figure 5 shown that the recovery ratio of I_{dsat} is reduced with higher drain bias($|V_{ds}|>2V$). ΔV_{th} and I_{dsat} degradation have the linear relationship after stress [3] and ΔV_{th} after recovery can be calculated by ΔI_{dsat}. In the conventional charge pumping measurement, the interface trap has large recovery [7] and there is some time delay between stress and CP measure. With smaller recovery ratio, more N_{it} can be swept by CP pulse. Figure 4 demonstrates that the relationship of ΔV_{th} after 500*s* recovery and ΔN_{it} degradation.

4. Conclusion

The frequency-dependent charge pumping techniques have been used to study the characteristics of Si-SiO_2 interface trap generation under DBNBTI stress. The results indicate that the creation of new interface traps with threshold degradation is similar to NBTI degradation under drain bias, which could be owned to the drain bias enhanced ≡Si-H bond breaking. The drain bias can enhance the generation of non-recovered interfacial traps and cause the smaller I_{dsat} recovery ratio.

Ex. 1016, Page 248

Acknowledgments

The work is partly support by 973 & 863 Programs..

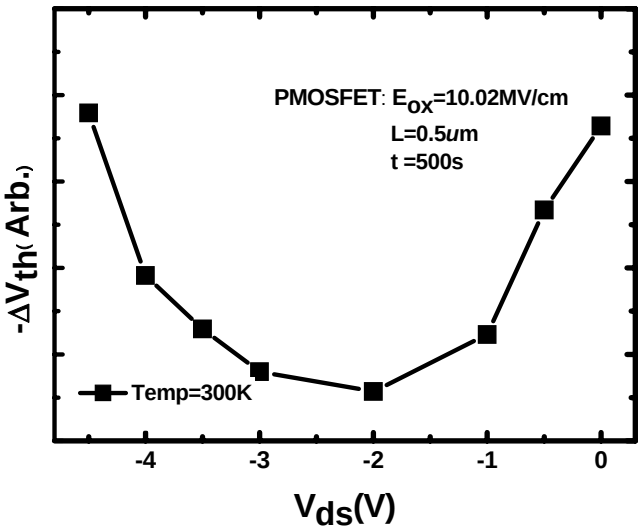


Figure 1. Shift of threshold voltage for 0.5*um* pMOSFET with applied drain bias, stress time=500*s*

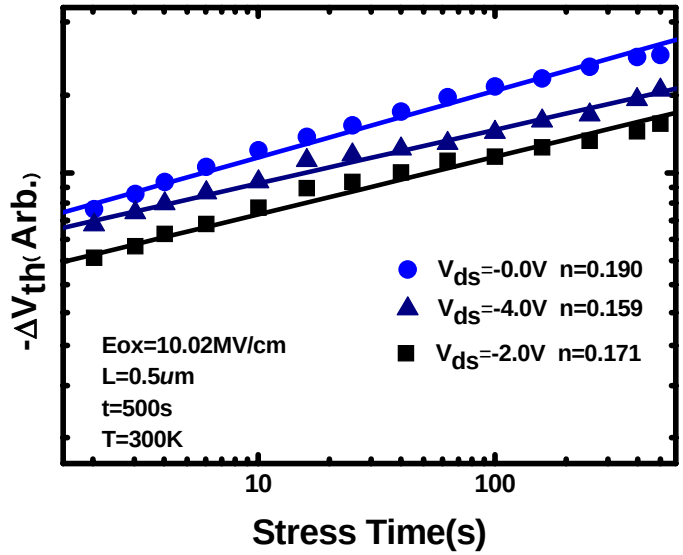


Figure 2. Threshold voltage degradation at E_{ox}=10.02*MV/cm* with different applied drain bias temperature=300*K*, the power law exponent vary between 0.155-0.195.

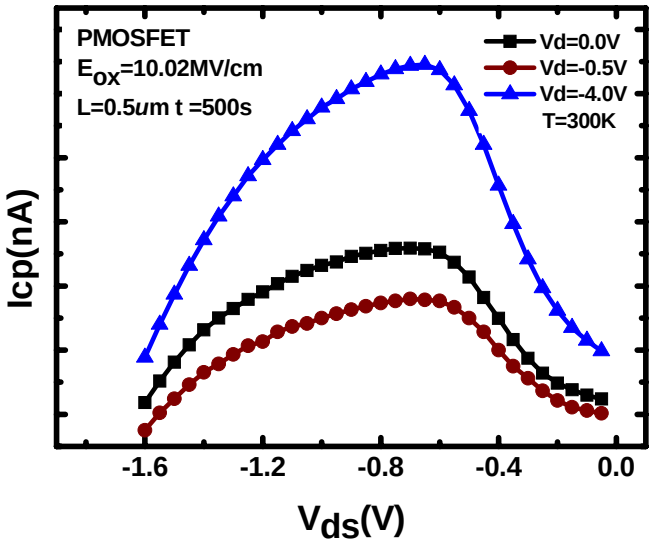


Figure 3. The 1MHz variable base-level charge pumping results of 0.5*um* pMOSFET under stand DBNBTI stress at temperature=300*K* and stress time=500*s*，pulse amplitude= 2*V*

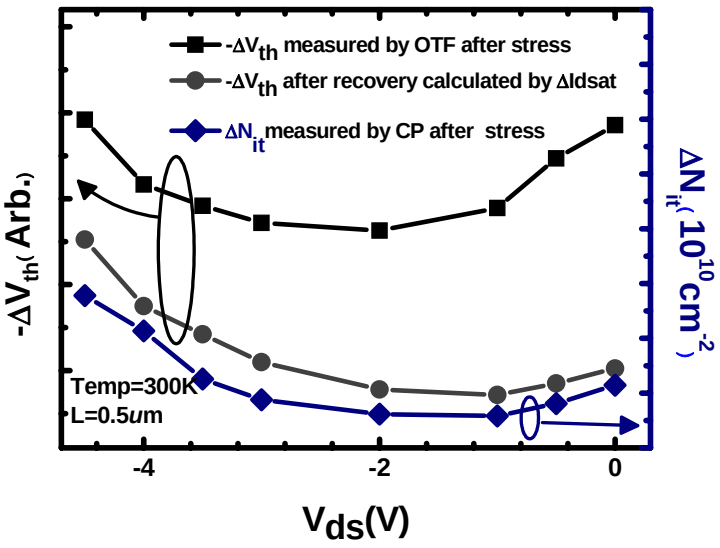


Figure 4. Shift of threshold voltage and generation of interface trap for 0.5*um* pMOSFET with applied drain bias at 300*k* and stress time=500*s*, compare with the calculated ΔV_{th} by ΔI_{dsat} after 500*s* recovery time, all have the turnaround phenomenon. Stress: $V_{gstress}$=-5.5*V*, Recovery: V_g=V_{ds}=0.0*V*

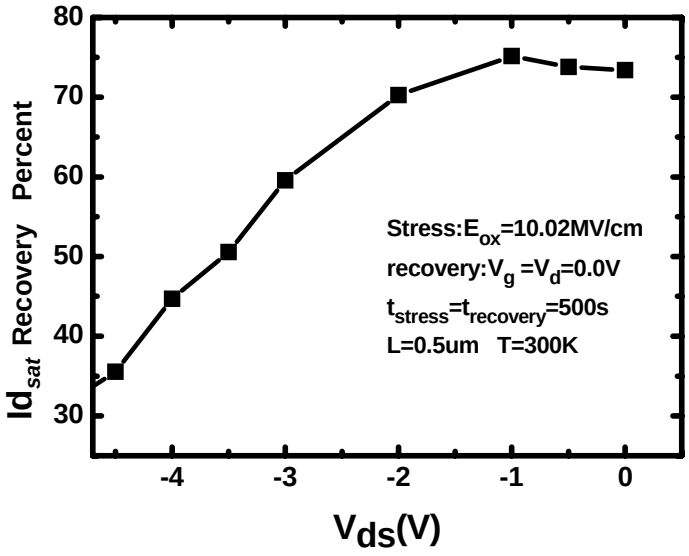


Figure 5. V_{ds} dependence of I_{dsat} recovery ratio, with high drain bias, the value is reduced after $|V_{ds}|>2V$. $V_{th\_initial}$ is around -0.7*V*, I_{dsat} was measured at $V_{g=}$ V_{ds} =-1.0*V*.

References

1. J.Q. Yang et al, *Proc. Solid State Devices and Material*, pp.350-351(2008).
2. N.K. Jha et al,, *Proc. International Reliability Physics Symposium* (*IRPS*), pp. 524 – 528 (2005).
3. Y.h. Luo et al, *Proc. International Reliability Physics Symposium* (*IRPS*), pp. 264-267 (2007).
4. Steve S.Chung et al, *Proc. Electron Devices and Semiconductor Technology (EDST)*, pp. 46-50 (2007).
5. M. Denais et al., *IEDM Tech. Dig.*, pp. 109 – 112 (2004).
6. S. Mahapatra et al, *IEDM Tech. Dig.*, pp.105-108 (2004).
7. W.J. Liu et al, Proc. International Reliability Physics Symposium (IRPS), pp. 813-816(2007).

Ex. 1016, Page 251

ECS Transactions, 18 (1) 225-229 (2009)
10.1149/1.3096454 © The Electrochemical Society

Challenges of Contacting Lead-Free Packages

Gerhard Gschwendtberger

Multitest elektronische Systeme GmbH, Rosenheim, Germany

Lead-free packages require dedicated contact spring coatings. In many cases these coatings can cause significant premature wear issues on the IC test socket which cause dramatic cost of test increases. Other coating solutions suffer from material transfer, which causes unrepeatable contact resistance readings and frequent test socket cleaning cycles. Again, the effect is a significant cost increase by test cell downtime and short socket lifespan. Test socket related cost of test numbers are significantly influenced by above described effects. Dependencies between test socket cleaning cycles, yield loss, test socket spare part consumption and cost of test numbers can be positively influenced by an optimized set-up of important parameters i.e. contact force, coating material, contact scrub, surface quality, contact geometry.

Impact of Lead-Free Packages onto Test Socket Performance

Lead-free packages and materials

The big challenge is the wide variety of currently used lead-free materials featuring heterogeneous parameters in terms of hardness, thickness, surface structure, oxides etc. Typically we find Matte Sn, SnBi, SnAg and SnAgCu for SOP, QFN, QFP and BGA packages. But for SOP, QFP and QFN packages we also often see PdNi, PdNiAu. In addition several others materials and compositions are used.

The following considerations focus on PdNiAu and Matte Sn.

Example for heterogeneous material properties.

Lead-free Materials	PdNiAu	Matte SN
Plating Thickness	1µm – 2 µm	10 µm – 20 µm
Plating Hardness	300 HV	20 – 30 HV

Effects of lead-free materials onto test sockets

Based on the different properties the materials require adequate contacts concepts and contact spring coatings to avoid negative impacts on the contacting performance. E.g. the low hardness of Matte Sn causes material transfer on the contact springs, which finally badly influence the repeatability of the contact resistance. On the other hand NiPdAu is very hard and therefore will lead to a premature wear out of the contact spring – again the repeatability of the contact resistance is deteriorating.

Ex. 1016, Page 253

Also the contact concepts of the sockets are subjected to the materials as shown in figure 1.

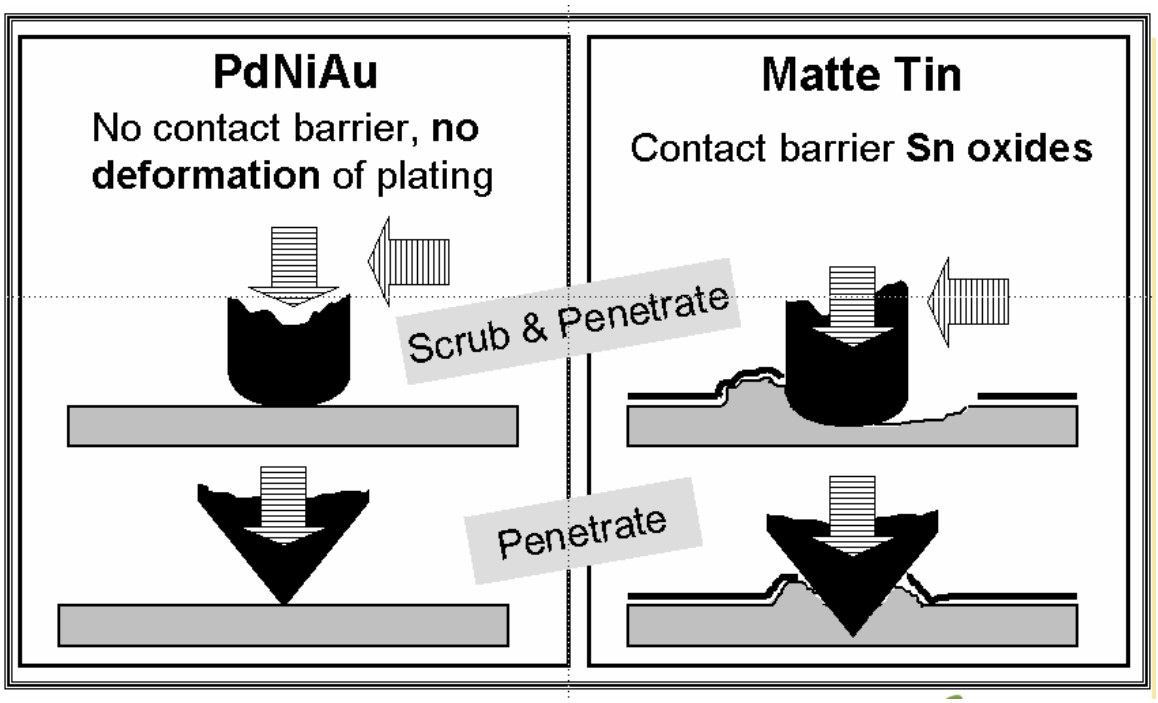


Figure 1. Contact concepts vs. Pb-free materials.

Impact of Test Socket Performance on Cost of Test

Although the test socket itself is only a very small part of the test cell and only requires small investment compared to test head and test handler etc., it can have significant impact in the overall test cell OEE (overall equipment efficiency). The following section will show how core OEE parameters - test cell yield and test cell availability – are influenced by test socket cleaning cycles, reduced lifespan of contact spring and unstable contact resistance. All examples are calculated per year for an application with 1 second test time and a 365 days/24 hours utilization.

Test socket cleaning cycles

Each test socket cleaning will lead to a test cell down time. The following diagram (figure 3.) relates the test cell down times caused by test socket cleaning to the capital depreciation cost of the test cell (1 Mio Euro investment, 5 year depreciation period).

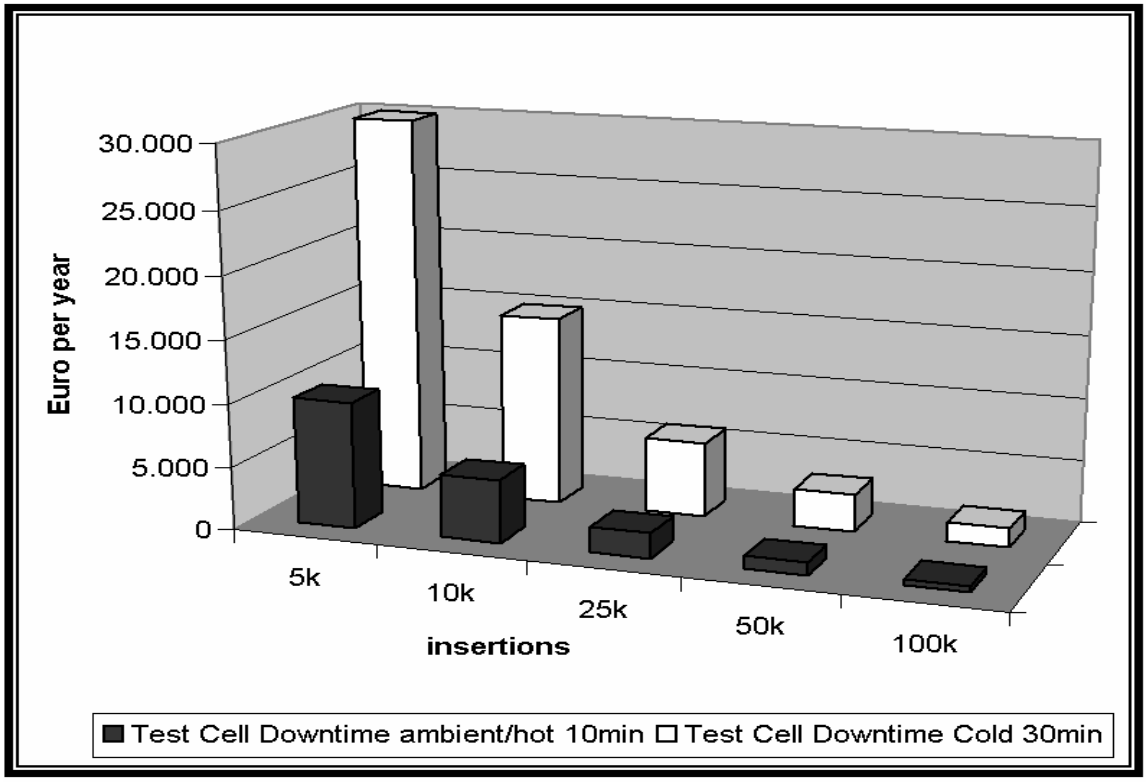


Figure 2. Cleaning cycles in number of insertions vs. depreciation costs per year

Test socket lifespan

Premature wear out and shortened lifespan have a direct influence on cost, as they require either spare parts or complete replacements of the sockets. The following example shows the spare part cost for one year at 70% OEE and a spare price of 200 Euro.

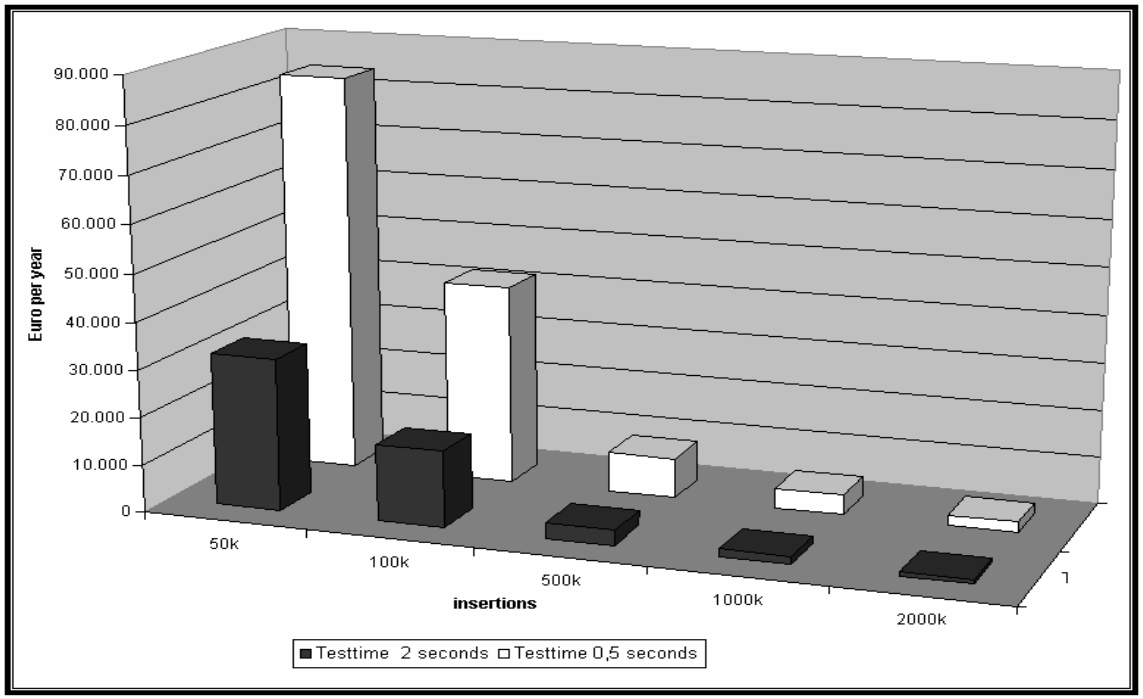


Figure 3. Test socket lifespan (insertions) vs. spare parts cost per year

Test socket related yield loss

Worn out or contaminated contact springs will cause bad test yield. Figure 4. shows the loss in Euro for an IC which has a price of 0.1 Euro in relation to a reduction of the overall test yield caused by deteriorating contacting performance.

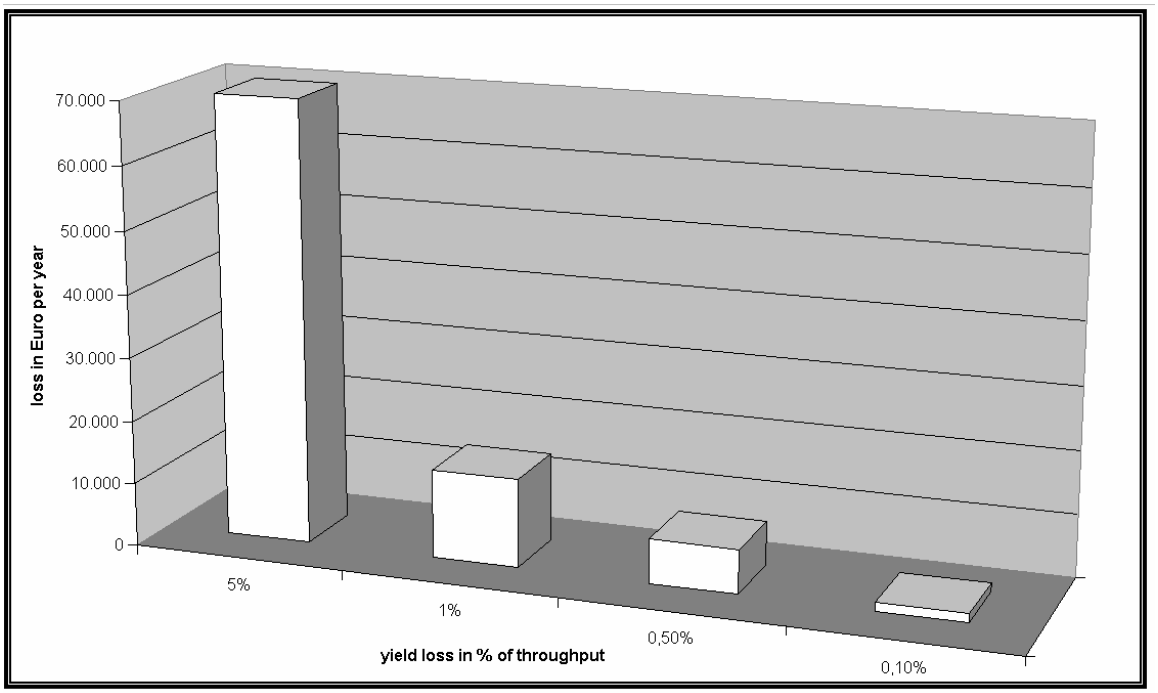


Figure 4. Test socket lifespan (insertions) vs. spare parts cost per year

Optimizing Test Socket Parameters for Lead-Free Packages

Pb-free requirements for test sockets

Lead-free packages need dedicated contacting solutions in order to ensure proper test performance:

- Contact spring wear resistance against PdNiAu
- No or minimum Sn transfer and accumulation
- Remove / protrude oxide barriers of Sn
- Minimum impact onto IC lead plating
- Compatible electrical behavior, contact resistance, current capability, signal fidelity etc.

Test socket features and characterization

There is a set of test socket features, which directly responds to these needs:

- Contact force

Ex. 1016, Page 256

- Plating
- Scrub
- Geometry
- Surface

In order to ensure real and consistent performance improvement, all of these parameters need to be reviewed, optimized and synchronized against each other. Experience shows that this cannot be done in a classical R&D situation. As field studies will take to long and will unacceptably the time-to-market, a procedure and set-up is needed that allows for test socket characterization in a situation as close as possible to real high volume production environment.

Ex. 1016, Page 257

ECS Transactions, 18 (1) 231-236 (2009)
10.1149/1.3096455 © The Electrochemical Society

POWER-AWARE TEST GENERATION FOR REDUCING YIELD LOSS RISK IN AT-SPEED SCAN TESTING

Y. Yamato[a], X. Wen[a], K. Miyase[a], H. Furukawa[a], and S. Kajihara[a]

[a] Kyushu Institute of Technology, Iizuka 820-8502, Japan

Power-aware *X*-filling is a preferable approach to avoiding IR-drop-induced yield loss in at-speed scan testing without the need of any circuit modification. However, the effect achieved by previous *X*-filling methods for reducing launch switching activity may be far from optimal. In addition, some of them are not scalable. This paper proposes a new *X*-filling method based on the genetic algorithm (GA). Experimental results on benchmark circuits demonstrate the effectiveness and scalability of the new *X*-filling method for reducing launch switching activity.

1. INTRODUCTION

With shrinking feature sizes, growing clock frequency, and decreasing power supply voltage, modern integrated circuits (IC's) are increasingly suffering from timing related defects, usually of small delays [1]. As a result, delay testing has become mandatory in order to reduce the defect level of deep submicron (DSM) circuits.

Since most of circuits employ scan design, ***at-speed scan testing*** is widely used for delay testing in practice [1]. Fig. 1 shows the timing diagram of at-speed scan testing based on the ***launch-on-capture*** (***LOC***) scheme [1]. In ***shift*** mode (SE = 1), a test vector is applied by operating scan chains as one or more shift registers over multiple shift clock pulses, with S_L as the last shift clock pulse. In ***capture*** mode (SE = 0), two capture pulses are applied: C_1 (***launch capture***) for launching a transition at the start-point of a path and C_2 (***response capture***) for capturing the circuit response to the launched transition at the end-point of the path. Note that the ***test cycle*** is equal to the functional clock cycle, which is very short for a high-speed design, in at-speed scan testing.

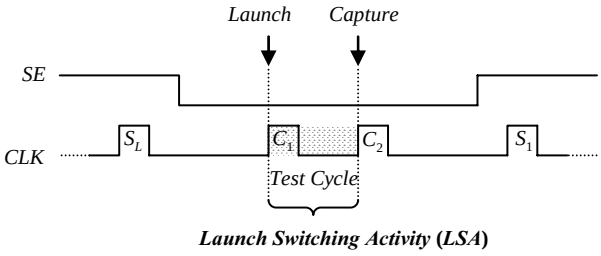


Fig. 1 Launch-on-capture Scheme

Switching activity due to the launch capture C_1 for a test vector, referred to as ***launch switching activity*** (***LSA***), is usually higher (33% higher as reported in [2]) than that of a functional vector. This may lead to excessive ***IR-drop*** in the power distribution network, increasing gate delays along paths in the test cycle so that timing violations may occur at the response capture C_2 only during at-speed scan testing. As a result, undue costly yield loss may occur, which is unacceptable. This yield loss problem is especially severe for high-speed circuits, whose test cycles are short. Low-power designs are also vulnerable due to narrower noise margins because of reduced power-supply voltages. For this reason, at-speed scan testing is extremely sensitive to IR-drop. Therefore, it is highly important to reduce launch switching activity during at-speed scan testing.

Ex. 1016, Page 259

Previous techniques for reducing launch switching activity can be classified into two categories. ***Circuit- modification-based*** techniques include *additional circuitry insertion* [3, 4], *scan chain segmentation* [5], *scan re-ordering* [6], and *partial capturing* [7], while ***data-manipulation-based*** techniques include *test vector reordering* [8], *test generation* [9, 18], and *X-filling* [10-14]. Generally, the approach of data manipulation is preferable, since it avoids the cost of circuit overhead and the risk of performance degradation due to the approach of circuit modification. Especially, ***X-filling*** is now widely used in practice for reducing launch switching activity, since it is compatible with any existing ATPG flow, and has no impact on test data volume, area overhead, and circuit timing.
X-filling is the process of properly assigning logic values to the *X*-bits in a test cube for achieving a specific goal. Here, the goal is to reduce launch switching activity in at-speed scan testing. This is usually achieved by minimizing either the Hamming distance between a test vector and its response at FFs [10-13] or the ***weighted switching activity*** (***WSA***) for all nodes in a circuit [14].

Until now, three basic approaches have been used for *X*-filling, namely ***justification-based***, ***probability- based***, and ***hybrid***.

The method in [11] is a typical justification-based technique. PPI-PPO bit-pairs of the form <logic value, *X*> are processed first, followed by those of the form <*X*, *X*>, one at a time. In case of bit-pairs of the form <logic value, *X*>, the PPO *X*-bit is justified to the logic value of its corresponding PPI. Then, bit-pairs of the form <*X*, *X*> are processed. Note that one such bit-pair may require two passes of justification. For example, for $<r_1, r_2> = <X, X>$, 1 is first justified on r_2 and 1 is assigned to r_1. If this fails, 0 is then justified on r_2 and 0 is assigned to r_1. This approach is highly effective. However, it is less scalable due to long run time since only one pair of bit-pair is processed at a time.

The methods in [12, 14] are typical probability-based techniques. The 0 and 1 probabilities of each node are first calculated by setting 0.50 as the 0 and 1 probabilities for each input *X*-bit and conducting probability propagation. Then, the logic value for a PPI *X*-bit is determined by using the relation between the 0 and 1 probabilities of its corresponding PPO *X*-bit. In the method in [12], once 0 and 1 probabilities are calculated, all PPI *X*-bits in a test cube are assigned at one time to achieve high scalability. In contrast, in the method in [14], 0 and 1 probabilities are iteratively calculated with each *X*-bit assignment to achieve high effectiveness. However, the effect of such probability-based technique may be damaged by inevitable approximation in probability calculation. If the difference between the 0 and 1 probabilities at a PPO *X*-bit is negligibly small, logic value determination for its corresponding PPI *X*-bit becomes highly inaccurate. Moreover, reconverging paths may cause significant errors in probability calculation.

The method in [13] is a typical hybrid technique employing two operations, namely ***limited justification*** and ***probability recalculation***. The first operation is conducted for PPI-PPO bit-pairs of the form <logic value, *X*>, but not for any PPI-PPO bit-pair of the form <*X*, *X*> for which multiple passes of justification may be needed. This is to achieve higher scalability. The second operation is conducted for PPI-PPO bit-pairs of the form <*X*, *X*>, but in multiple passes. That is, the logic value for a PPI *X*-bit is determined only if its corresponding PPO *X*-bit has significantly different 0 and 1 probabilities; otherwise, probabilities are re-calculated in the next pass. This is to achieve higher effectiveness through improved accuracy.

Recently, methods to reduce both shift and capture power are proposed [19, 20]. Especially [19] can reduce not only shift-in power but also shift-out power.

Ex. 1016, Page 260

However, the effect achieved by previous *X*-filling methods for reducing launch switching activity in at-speed scan testing may be far from optimal. In addition, some of them are not scalable. In order to achieve both high effectiveness and scalability, we propose a new *X*-filling method based on the ***genetic algorithm*** (GA).

2. PROPOSED METHOD

2.1 Background

GA is a search technique used in computing to find exact or approximate solutions to optimization and search problems [16, 17]. Such an algorithm is started with a set of solutions (***individuals***), represented by ***chromosomes***, called ***population***. ***Fitness*** is then calculated for each individual in the population. Individuals are taken and used to breed a new population according to their fitness. This is repeated until a certain preset termination condition, e.g. the number of populations or improvement of the best individual, is satisfied.

Conceptually, GA is an ideal fit for the problem of *X*-filling for reducing launch switching activity in at-speed testing. This is because finding a good solution for *X*-filling needs a search process through a large solution space.

2.2 General Flow

The general flow of the proposed GA-based *X*-filling method, called ***GA-fill***, is shown in Fig. 2. Its details are described in the following.

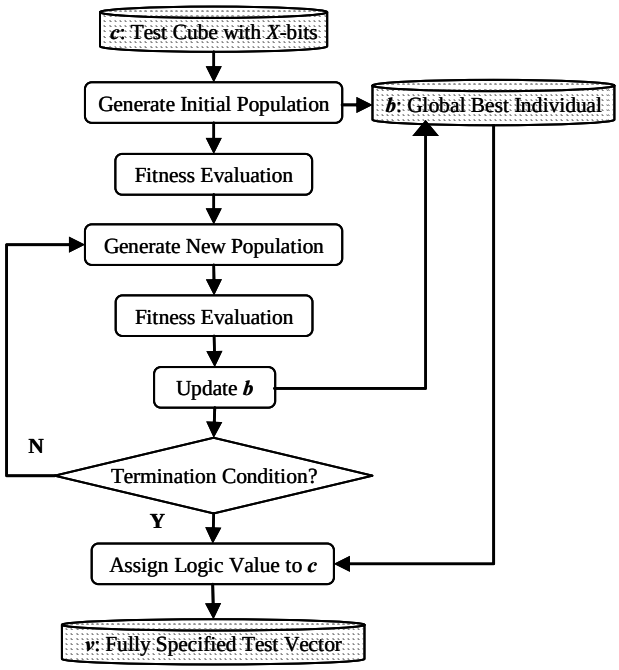


Fig. 2 General Flow of GA-fill

In GA-fill, a set of *X*-bits in a test cube is considered as an individual. First, an initial population including a certain number of individuals is randomly generated, and then each chromosome (0 or 1) in an individual is assigned to corresponding *X*-bits and logic simulation is conducted to collect information necessary for evaluating the fitness of the

individuals. Here, pattern- parallel logic simulation for multiple individuals is used in order to reduce computation time.

The following evaluation function, denoted by *fitness*(*i*), is used to calculate the fitness of each individual *i*:

$$fitness(i) = 1 - \frac{WSA\_FF(i)}{\#Scan\_FF} \quad \text{..... (i)}$$

if the ratio of *X*-bits in a test cube *c* is over 60%; otherwise,

$$fitness(i) = 1 - \frac{WSA\_Node(i)}{\#Node} \quad \text{..... (ii)}$$

Here, *WSA_FF*(*i*) and *#Scan_FF* in expression (i) are the WSA value for all scan FFs with respect to the assignment in *i* and the number of Scan FFs in a circuit, respectively. *WSA_Node*(*i*) and *#Node* in expression (ii) are the WSA value for all nodes with respect to the assignment in *i* and the number of nodes in a circuit, respectively. The computation time of expression (ii) is generally longer than that of expression (i), but the use of expression (ii) is obviously more effective for the purpose of reducing WSA for all nodes in a circuit. GA-fill selectively uses these two evaluation functions in order to achieve both high effectiveness and good scalability.

After the initialization, new populations are iteratively generated. During each successive generation, a pair of parent individuals is selected to breed a child individual. Individuals are selected through a certain criteria, where fitter individuals are typically more likely to be selected. The next step is to generate a child from those selected parents through genetic operators: ***crossover*** and/or ***mutation***. A child is created which typically shares many of the characteristics of its parents, and chromosomes in the child are flipped with low probability in mutation to keep the diversity of individuals. New parents are selected for each child, and the process continues until a new population of individuals of an appropriate size is generated. Note that an individual with the highest fitness throughout the whole population is kept. This process is repeated until certain termination conditions have been reached. As a result, logic values are assigned to *X*-bits in the test cube based on the globally fittest individual.

2.3 Advantages

GA-fill has two major advantages. One, as stated above, is that GA-fill tends to achieve a more optimal solution in a more efficient manner. Another is that GA-fill is adaptive for achieving any specific optimization object. In this paper, the object is set to be the reduction of the launch switching activity measured by the WSA metric. This is achieved with a proper evaluation function, such as the one of (i) and (ii).

However, reducing the WSA in a whole circuit may be insufficient to avoid test-induced yield loss. For this reason, the method in [21] focuses on reducing the launch switching activity on gates that are close to activated critical paths. GA-fill is especially suitable for such fine-grain reduction of launch switching activity, by simply using a proper evaluation function.

3. EXPERIMENTAL RESULTS

GA-Fill was implemented in C programming language, and experiments were conducted on five largest ISCAS'89 benchmark circuits using a computer with 3.0GHz

CPU and 24GB memory. Test vectors were generated by using an internal ATPG tool and test cubes were then obtained from them by using an internal test relaxation tool [15]. The transition delay fault model was used. The roulette wheel selection and two-point crossover were employed for parent selection and crossover, respectively, while mutation probability was set to 0.1% in the experiments.

Table 1 summarizes the *X*-filling results by various methods: justification-based method [11] with high effectiveness, probability-based method [12] with high scalability, and the proposed GA-fill method. The ***WSA*** (***Weighted Switching Activity***) metric [6, 12] was used for estimating the LSA. Here, the weight of a node was set to be the number of its fanouts plus 1. If a node has no fanout, its weight was set to 1.

In Table 1, the test set information of each circuit, including the test vector count ("*Test Vec. #*"), fault coverage ("*Fault Cov.* (%)"), percentage of identified and *X*-bits ("*X* (%)"), is shown. In addition, maximum WSA reduction ratio at all nodes ("*Max. WSA Reduction Ratio* (%)") and the CPU time ("*CPU* (s)") are provided.

Table 1 Results of Various *X*-Filling Methods

Circuit	Test Set			Max. WSA Reduction Ratio (%)			CPU (s)		
	Test Vec. #	Fault Cov. (%)	*X* (%)	[11]	[12]	**GA**	[11]	[12]	**GA**
s13207	323	79.5	67.4	11.7	7.9	**13.2**	40	9	**28**
s15850	221	70.2	54.3	26.2	23.3	**27.6**	33	8	**29**
s35932	337	82.5	92.0	7.6	10.0	**10.4**	230	26	**152**
s38417	270	98.0	47.1	14.2	14.2	**20.1**	654	27	**296**
s38584	412	83.9	73.9	14.9	8.7	**15.0**	2086	42	**194**
Ave.	—	—	54.7	14.9	12.8	**17.3**	—	—	—

As shown in Table 1, the GA-fill method achieved the best results with respect to the reduction ratio for the maximum WSA at all nodes in a circuit. In addition, for scalability, the result of the GA-fill method was much closer to that of the best-performing method [12] than the justification-based method [11]. This shows that the GA-fill method can achieve both higher effectiveness and scalability than previous methods.

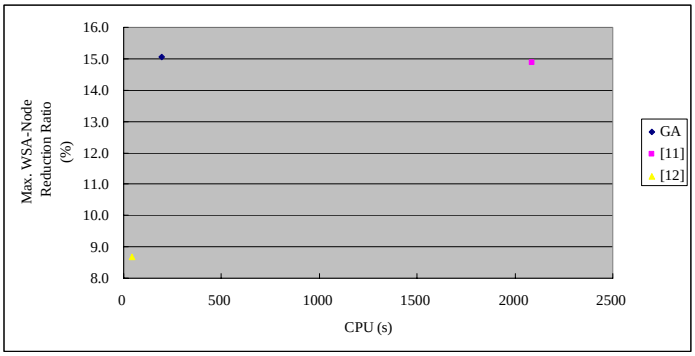


Fig. 3 CPU Time vs. WSA Reduction for s38584

The relation between the CPU time and the maximum WSA reduction ratio at all nodes of each method on s38584 is shown in Fig. 3. This result further demonstrates that the GA-fill method can achieve both effectiveness and scalability in a more balanced manner than previous methods.

4. CONCLUSIONS

This paper proposed a novel *X*-filling method, called *GA-fill*, which reduces the launch switching activity in at-speed scan testing with high effectiveness and scalability. Generally, GA-fill has a better chance to achieve optimal or near-optimal solutions in a reasonable amount of time, even though good solutions are hard to find for the problem by any deterministic method. Experimental results have confirmed the capability of the proposed method.

Future work includes (1) fine-tuning GA-fill parameters, especially the evaluation function, in order to further improve its effectiveness and scalability, (2) conducting further experiments using more circuits, including industrial circuits, (3) applying the GA-fill method for fine-grain reduction of launch switching activity, and (4) extending the current GA-fill method to the test compression environment.

REFERENCES

[1] L.-T. Wang, C.W. Wu, and X. Wen, (Editors), *VLSI Test Principles and Architectures: Design for Testability*, Elsevier, 2006.

[2] J. Saxena, K. Butler, V. B. Jayaram, and S. Kundu, "A Case Study of IR-Drop in Structured At-Speed Testing," *Proc. ITC*, pp. 1098-1104, 2003.

[3] R. Sankaralingam and N. A. Touba, "Inserting Test Points to Control Peak Power During Scan Testing", *Proc. Intl. Symp. on DFT*, pp. 138-146, 2002.

[4] S. Gerstendoerfer and H. Wunderlich, "Minimized Power Consumption for Scan-Based BIST," *Proc. ITC*, pp. 77-84, 1999.

[5] K. Lee, T. Huang, and J. Chen, "Peak-Power Reduction for Multiple-Scan Circuits during Test Application," *Proc. ATS*, pp. 453-458, 2000.

[6] P. Girad, "Survey of Low-Power Testing of VLSI Circuits," *IEEE Design & Test of Computers*, Vol. 19, No. 3, pp. 82-92, 2002.

[7] S. Wang and W. Wei, "A Technique to Reduce Peak Current and Average Power Dissipation in Scan Designs by Limited Capture," *Proc. ASP-DAC*, pp. 810-816, 2007.

[8] V. Dabholkar, S. Chakravarty, I. Pomeranz, and S. Reddy, "Techniques for Minimizing Power Dissipation in Scan and Combinational Circuits during Test Application," *IEEE Trans. on CAD*, Vol. 17, No. 12, pp. 1325-1333, 1998.

[9] F. Corno, P. Prinetto, M. Redaudengo, and M. Reorda, "A Test Pattern Generation Methodology for Low Power Consumption," *Proc. VTS*, pp. 35-40, 1998.

[10] W. Li, S. M. Reddy, I. Pomeranz, "On Reducing Peak Current and Power during Test," *Proc. ISVLSI*, pp. 156-161, 2005.

[11] X. Wen, Y. Yamashita, S. Morishima, S. Kajihara, L.-T. Wang, K. K. Saluja, and K. Kinoshita, "Low-Capture-Power Test Generation for Scan-Based At-Speed Testing," *Proc. ITC*, Paper 39.2, 2005.

[12] S. Remersaro, X. Lin, Z. Zhang, S. M. Reddy, I. Pomeranz, and J. Rajski, "Preferred Fill: A Scalable Method to Reduce Capture Power for Scan Based Designs," *Proc. ITC*, Paper 32.2, 2006.

[13] X. Wen, K. Miyase, S. Kajihara, T. Suzuki, Y. Yamato, P. Girard, Y. Ohsumi, and L.-T. Wang, "A Novel Scheme to Reduce Power Supply Noise for High-Quality At-Speed Scan Testing," *Proc. Int'l Test Conf.*, Paper 25.1, 2007.

[14] X. Wen, K. Miyase, T. Suzuki, Y. Yamato, S. Kajihara, L.-T. Wang, and K. K. Saluja, "A Highly-Guided *X*-Filling Method for Effective Low-Capture-Power Scan Test Generation," *Proc. ICCD*, pp. 251-258, 2006.

[15] K. Miyase and S. Kajihara, "XID: Don't Care Identification of Test Patterns for Combinational Circuits," *IEEE Trans. on Computer-Aided Design*, Vol. 23, No. 2, pp. 321-326, Feb. 2004.

[16] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.

[17] J. Holland, *Adaptation in Natural and Artificial Systems*, The University of Michigan, 1975., and MIT Press, 1992.

[18] N. Ahmed, M. Tehranipoor, and V. Jayaram, "Supply Voltage Noise Aware ATPG for Transition Delay Faults," *Proc. VLSI Test Symp*., pp. 179–184. 2007,

[19] J. Li, Q. Xu, Y. Hu and X. Li, "iFill: An Impact-Oriented *X*-Filling Method for Shift- and Capture-Power Reduction in At-Speed Scan-Based Testing," *Proc. DATE*, pp. 1184-1189, 2008.

[20] S. Remersaro, X. Lin, S. M. Reddy, I. Pomeranz, J. Rajski, "Low Shift and Capture Power Scan Tests," *Proc. VLSI Design*, pp. 793-798, 2007.

[21] X. Wen, K. Miyase, T. Suzuki, S. Kajihara, Y. Ohsumi and K. K. Saluja, "Critical-Path-Aware *X*-Filling for Effective IR-Drop Reduction in At-Speed Scan Testing," *Proc. DAC*, pp. 527-532, 2007.

Ex. 1016, Page 264

ECS Transactions, 18 (1) 237-241 (2009)
10.1149/1.3096456 © The Electrochemical Society

Yield Learning Through Scan Diagnosis

Yu Huang Wu-Tung Cheng

Mentor Graphics Corp., 8005 S.W. Boeckman Road, Wilsonville, OR 97070, USA

Abstract

Increasing design complexities conspire to threaten long held assumptions about the predictability of design closure. Yield learning and ramp up become a very important issue in industry. High volume scan-based manufacturing diagnosis and data mining can be exploited as a capable contributor to identify systematic defects that have been the most important yield limiter. In this paper, this methodology will be illustrated by a few case studies.

Introduction

Technology developments in IC design and manufacturing lead to highly sophisticated lithography techniques and new materials that are challenging the controlled yield ramp-up of the new process generations below 100 nm feature size (1). With new materials and processes, the devices manufactured at nanometer levels tend to have different types of defects that can be classified into three categories: *random defects*, *systematic defects* and *parametric defects*. *Random defects* are independent of designs and processes. For example, random defects might be caused by random particles scattered on a wafer during manufacturing, impurities in substances used for manufacturing, missing or extra material on some layers, and equipment failures. *Systematic defects* depend on designs and processes used for manufacturing. For example, *systematic defects* might be caused by limitations of current EDA design tools, DFM rule violations and process variations. *Parametric defects* mean all elements of a chip are functioning, but timing or other electrical requirements, such as power, are not to specification. *Parametric defects* are the result of interconnect parasitics and device physics. They can cause a circuit to behave in non-ideal ways, such as causing power bus voltage drop, increased noise, timing skew in clock and signal lines.

As prior study pointed out, with the continuous shrinking of feature sizes, somewhere below the 180 nm technology node, *systematic defects* have a larger impact on yield than *random defects* (2). Therefore, in modern VLSI manufacturing, the goal of a yield monitoring process is primarily to identify systematic yield detractors and sensitive design features such that yield can be dramatically improved. Yield analysis has traditionally been a tedious manual process of analyzing wafers with lower-than-expected yield, identifying a failure signature, locating the defect source and finally discovering failure mechanism. Traditional techniques such as in-line inspection and memory bitmapping are becoming more inefficient as technologies become denser, and the number of metal layers increases.

Scan-based testing is a promising solution for maintaining high quality and low cost test for SOC designs because the scan chains embedded in the design can provide

Ex. 1016, Page 265

tremendous control and observation points. When introducing new or working on existing technologies, there is an increasing demand to analyze devices failed scan-based testing from the test floor and obtain more information about the defects such that faster and more reliable yield ramp-up becomes feasible. Therefore, in addition to an optimized pass/fail test, failure information from scan-based production test needs to be collected, diagnosed and analyzed. Compared to analysis approaches using either functional vectors or other traditional manual approaches, scan-based diagnosis takes in existing design information and ATE failure data and reports potential failure locations with possible defect types. It does not require any detailed low level design knowledge or long time frame to identify potential defective locations. It can guide physical failure analysis to identify the root cause of defects and can greatly minimize failure analysis cycle and reduce yield efforts (3). In the next section, several real case studies that had been published in recent years will be used to illustrate how scan-based diagnosis can be successfully applied in the process of yield improvement.

Case Studies

Case Study I

A yield learning flow based on layout-aware diagnosis was proposed in (4). The flow starts off with a given set of defect prone features extracted from layout. One example feature is two pointing corners of polygons in the same layer of the layout. Each feature implies one or more potential defects. In the case of the pointing corner the defect is a bridge. The test pattern set is transferred to an ATE and the test is executed. Here, it is important that the test does not stop on-first-fail, rather the failing data has to be collected for an adequate number of failing cycles. After diagnosis has been run on all fail logs, statistical analysis is started. The goal of statistical analysis is to compute the failure rates of all previously extracted features. Statistics can organize the data in many different ways, some of which don't directly relate to extracted features. For example, the overall defectivity of a metal layer or a particular library cell can be computed. For purposes of analysis, the statistical data can also be visualized, e.g. in the form of Pareto charts. It is possible that diagnosis cannot determine the reason for a failing die, thus these candidates will show up as undetermined. Further investigation is required in order to determine the root causes if the number of these cases becomes too large. Based on the computed failure rate, the yield limiting factors of the design can be determined. Depending on the type of the yield limiting factor, physical failure analysis may or may not be necessary to validate the finding. With this data a number of different applications are possible, which all base on the statistical results.

One of the most important applications is yield learning - using the statistical analysis results to identify yield limiting factors. Using real production data from LSI, Figure 1 plots the failure rate versus various types of layout features which are prone to bridging. As a reference the total number of each type of bridge in the design is also included in the chart. The X-axis shows all the bridge types and the Y-axis plots the (normalized) failure rates. The total number of bridge types in the design is given in a linear scale. The bridge types are abbreviated as S: side-to-side, C: corner-to-corner, E: end-of-line and V: via-to-via and W: side-to-side over wide metal. The digit next to the letter denotes the metal (via) layer the bridge is on.

Ex. 1016, Page 266

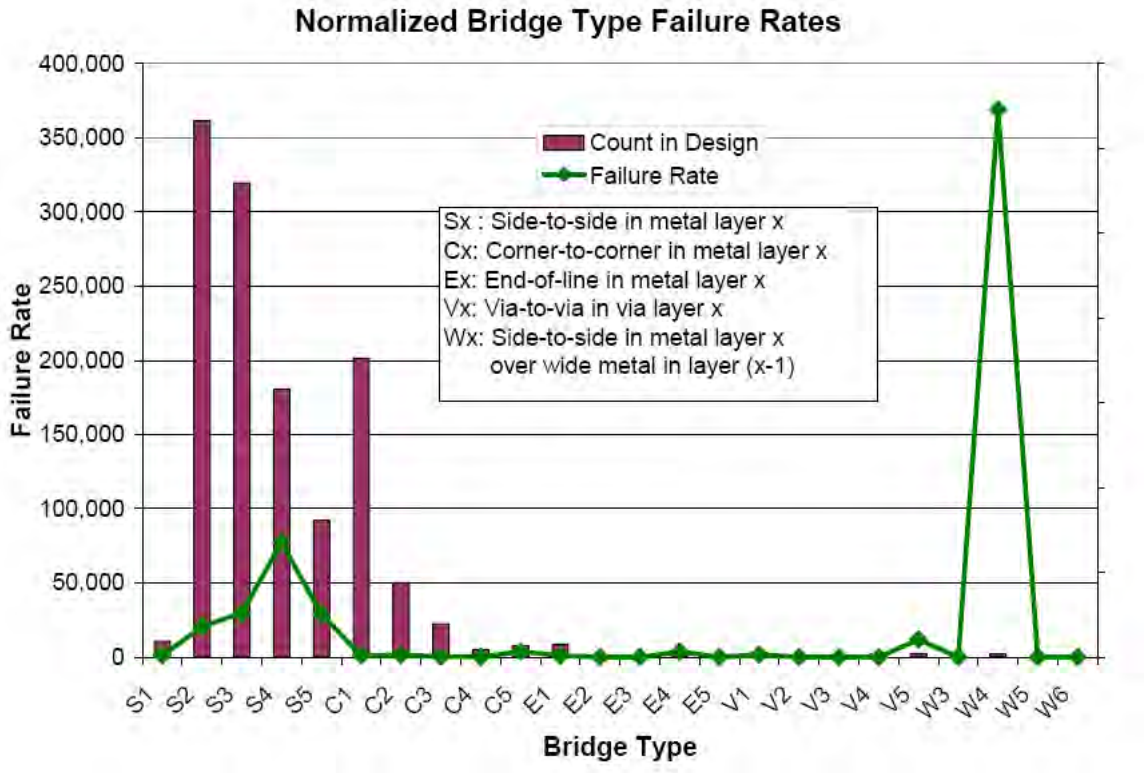


Figure 1. Normalized failure rate of all bridge types over all metal layers.

From Figure 1, it can be concluded that the side-to-side bridges over wide metal in metal layer 4 has a significantly higher failure rate than other kinds of bridges. For future design re-spins or future new designs, special consideration of avoiding such scenarios should be prioritized. In addition, ATPG should target this type of bridges more thoroughly than any other type of bridge. Furthermore, the fabrication process itself can be analyzed to determine what might be causing these types of bridges. This may eventually lead to modification of process steps to alleviate the issue.

Case Study II

In addition to the diagnosis tool, yield improvement through scan diagnosis is a team work which involves efforts from design engineers, test engineers, product engineers and failure analysis engineers. Figure 2 shows the diagnostics process flow adopted by Skyworks (5). One design manufactured at Skyworks had very low yield (about 20%) at the first silicon. A brief pre-screening of these ATE fail logs gave a hint that the majority of the failures were due to scan chain failures. By applying the adopted commercial diagnosis tool and flow, scan chains failures were reported. The diagnosis reports also identified the suspect scan cell IDs and reported they were hold-time faults.

In the design phases, timing constrain guard bands sometimes may not be sufficient due to process various or inaccurate timing models. Therefore, it is quite possible that a timing analysis tool didn't catch all timing related violation problems. After checking the static timing analysis report, it was confirmed that the reported suspect scan cells are likely had insufficient timing margin. In many cases like this one, a

Ex. 1016, Page 267

back end tool such as a scan diagnosis tool has to be used to find out a timing related issue during first silicon debug and yield ramp up stages.

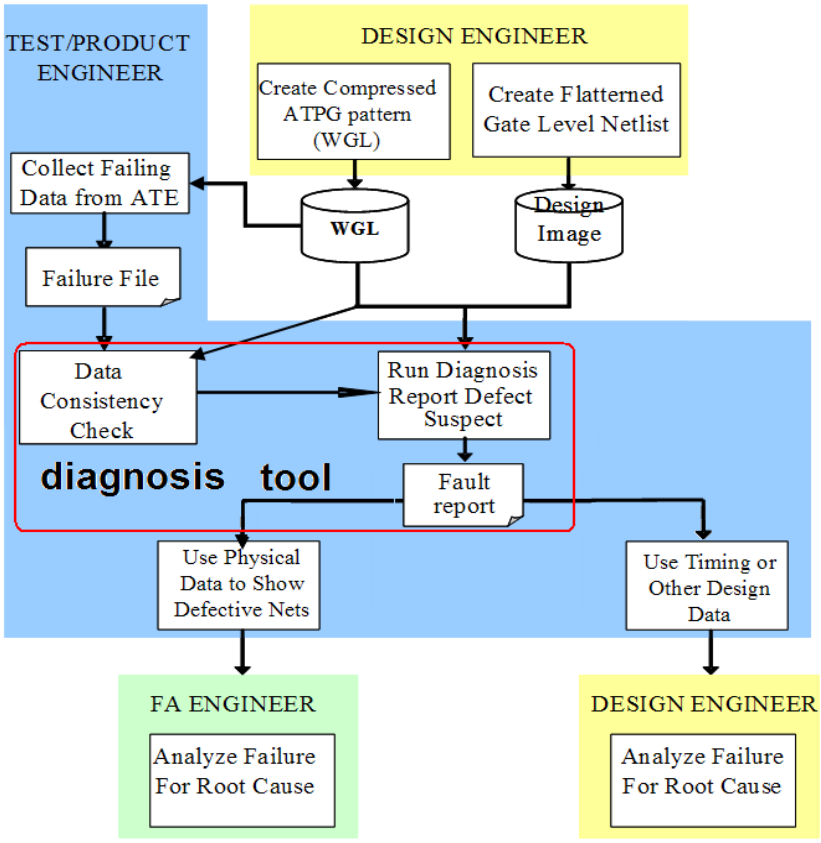


Figure 2. Diagnostics process flow at Skyworks.

In this design, the clock uncertainty used was 100ps in ATPG test mode. After identifying this systematic scan chain hold-time marginality, they fixed the marginal hold-time violations that existed on 72 flip-flops on 22 scan chains. In addition, they retimed the chip with clock uncertainty of 150 ps to make sure there are no marginality issues. After the fix and re-spin of the chip it has been seen that the yield ramped up from the original 20 to 90 percent.

Case Study III

Another yield improvement through scan-based diagnosis was reported in (6). Based on the die size, foundry's defect density for the technology and scan chain fault coverage, the expected fallout for scan chain test should be less than 1%, for a digital integrated circuit device still under development. However in this particular case, the data analysis from the first design tolerance split condition wafer lot revealed an abnormal 13% on average yield loss associated with only scan chain failures. These chain failures did not correlate with any particular wafer split conditions of the silicon processing or show any temperature, voltage or timing sensitivities.

Physical Failure Analysis (PFA) were performed on some diagnosed failing chips, and it showed that there was a resistive open defect on a metal layer due to poor copper

Ex. 1016, Page 268

fill. However, the defects seem to be random rather than systematic, in terms of cell types, defect types and physical locations etc. Normally, it is challenging to improve yield by reducing random defects. However, in this case, they had some interesting observations. The physical failure analysis result helped them to focus on backend layers for defect reduction. They implemented process steps for all the metal layers to reduce the defect count. The most prominent improvement was seen when they implemented additional drying step which included more spinning and dry air blowing after cleaning the wafer surface with distilled water.

In the normal process step, after cleaning the wafer surfaces with distilled water the wafers are spun to dry the surface. However in this particular case, the spinning was not sufficient enough to dry all the water droplets. There were some leftover water droplets that dissolve some photo resist when depositing the photo resist. Hence, in the following etching step, the metal below the dissolved photo resist spots gets over etched leading to an improper copper fill.

The leftover water droplets were present on the edge of the wafers due to the centrifugal force exerted while spinning the wafers, which correlated with the location of the majority scan chain failures in the wafer map. Wafer maps showed that there was higher percentage of scan chain failures in the edge of the wafer versus the center of the wafer. The exact fault location in the scan chain failing units was isolated with the use of Mentor's Yield Diagnostic Flow.

This process enhancement helped them to reduce the scan chain failures from 13% to 3%. It is an interesting experience that diagnosis and physical failure analysis on randomly distributed chain defects can still be helpful to enhance systematic process, and hence improve the yield. This is because although the defects seem to be random, they are in fact caused by systematic process issues.

References

1. A. Leininger, et. al., "Compression Mode Diagnosis Enables High Volume Monitoring Diagnosis Flow," *Proc. Int'l Test Conf.*, Paper 7.3, (2005).
2. R. Radojcic and M. Rencher, "Old Rules No Longer Apply," EETimes, April 29, (2003).
3. C. Eddleman et. al., "Advanced Scan Diagnosis Based Fault Isolation and Defect Identification for Yield Learning," ISTFA (2005).
4. M. Keim, et. al., "A Rapid Yield Learning Flow Based on Production Integrated Layout-Aware Diagnosis", *Proc. Int'l Test Conf.*, Paper 7.1, (2006).
5. I. Ahmed, et al., "Yield Improvement with Compressed Pattern Diagnosis," Proc. Int'l Workshop on Silicon Debug and Diagnosis, (2006).
6. G. Ponnuvel, et al., "Scan Chain Failure Diagnosis for Yield Improvement", Workshop on RTL and High Level Testing (WRTLT), (2008).

ECS Transactions, 18 (1) 243-248 (2009)
10.1149/1.3096457 © The Electrochemical Society

A Low Power Architecture based on Unidirectional Cut

Z. Chen[a], D. Xiang[b]

[a] Department of Computer Science and Technology, Tsinghua Universit, Beijing, 100084, China
[b] School of Software, Tsinghua Universit, Beijing, 100084, China

Test power of a circuit mainly consists of capture power and shift power. In this paper, a new scan architecture is proposed to reduce both of them. In the architecture, a new graph algorithm is proposed to partition the scan flip-flops into two parts and then the ones with common successors in each part are grouped as a chain. By partitioning the scan flip-flops into two parts, the process of capture is cut into two sequential steps. This technique can effectively reduce the capture power. Using the clock disabling scheme, we can make sure that, only one or a small portion of scan chains are active during the shifting phase. Therefore, the peak power and average power are reduced. This architecture can effectively reduce the test time too. Experimental result shows that average power and peak power reduction are very significant, compared with the ordinary full-scan architecture.

Introduction

Scan-based testing is widely used in modern VLSI design and test. Although scan testing simplifies ATPG, it introduces some serious problems: (1) test power may be much higher, (2) test data volume may be greatly increased, and (3) test application time may be much longer [1].

The power dissipation during scan-based testing may be much higher than that during the normal operation due to the excessive switching activities. These switching activities occur in the shift or capture operations. Several works have been proposed to reduce the switch activities. The methods to reduce shift power include test data compression, pattern modification [2], clock disabling [3, 4] and scan chain partition [4]. While the methods to reduce capture power include post-generation filling [2], modified test generation [5], scan chain partition and disabling [4].

The aforementioned methods mainly focus on reducing shift power or capture power separately. Recently, many methods are presented to reduce both shift power and capture power. They can be categorized into pattern based methods and DFT based methods.

In the scan architectures [4, 7], which can reduce both shift power and capture power, shift power is reduced by allowing only a part of scan chains to receive test data. Capture power is reduced, either by using a part of scan flip-flops to receive the response [4, 7], or by capturing response in different cycles [4].

Therefore, we try to use DFT based method to reduce power with little area overhead. In the previous DFT based methods [4, 7], which can reduce both shift and capture power, they use the correlation of patterns to partition the scan chains. These methods are test set dependent and the results depend on the character of test vectors. Different from them, we use the structure based method to partition, which is deterministic and efficient for a circuit, and use filling strategy to further reduce test power. Although the method in [4]

Ex. 1016, Page 271

also uses the structure based method, the area overhead may be prohibitive when the number of scan chains is increasing. Our method also can reduce test time simultaneously, and the experiment shows that the proposed method can achieve more power reduction than others.

In this paer, a new concept called unidirectional cut is proposed to partition the scan flip-flops into two parts, part1 and part2. In the partition result, there is no scan flip-flop in part2 can reach any scan flip-flop in part1. Based on this method, capture power can be reduced.

Our architecture addresses these issues mainly by using scan chain partition and disabling. Scan flip-flops are partitioned into two parts. The principle of scan flip-flop partitioning is that scan flip-flops in the first part are not reachable from any scan flip-flop in the second part(unidirectional cut). Therefore, test response related to scan flip-flops in the first part is independent of the test data to scan flip-flops in the second part. After partitioning the whole scan flip-flops into two parts, partition should be continued in each part. In each part, scan flip-flops with common successors are grouped together to construct a scan chain. Using this partitioning method, only one or some scan chains are needed to be activated for fault effect activation and fault effect propagation. And other scan chains can be disabled or provide more space for X-filling to reduce power.

For our method, it is easy to integrate into the existing ATPG flow by adding a post-processing step for extra vector of each pattern. For the new architecture and application scheme, the mainly hardware modification is adding the capacity to gate the clock of one partition.

The rest of this paper is organized as follows. The scan architecture is presented in Section 2. The algorithm and strategy are proposed in Section 3. Experimental results are presented in Section 4, followed by conclusions in section 5.

The Scan Architecture

In this section, the new scan architecture, which can reduce test power and test time, is introduced first as shown in Figure 1.

In the architecture, the scan flip-flops are first partitioned into two parts, part1 and part2, as described in Figure 1. And the partition principle is that no flip-flop of part2 can reach any flip-flop of part1. Based on this principle, the traditional capture operation, all flip-flops capturing in one cycle, can be cut into two sequential steps. The new capture operation is that the flip-flops in the part2 first capture their responses and the flip-flops in part1 capture their responses in the next cycle. Because no flip-flop of part2 can reach any flip-flop of part1, after the flip-flops in the part2 capturing their responses, the changed values of flip-flops in part2 have no impact on the responses of flip-flops in part1. The detailed partition algorithm is presented in section 3.

Then, scan flip-flops of part1 and part2 are partitioned into scan chains separately as Figure 1, and the partition principle is that flip-flops with common successors are grouped into one scan chain. Because the scan flip-flops that have common successors may be necessary to activate a fault and propagate the fault effect simultaneously, the shift power can be reduced for the reason that only a small portion of scan chains are activated during each clock cycle.

Clock disabling is implemented by using an extra pin C with a counter. The output of the counter is set to 1 for n consecutive clock cycles. Test data is shifted into the activated scan chains in part1 when C is set to 1, all scan chains in part2 are disabled during this period. The scan chains in part1 are disabled when C while scan chains in part2 are activated. Again, n consecutive clock cycles are used to shift the test data into the

Ex. 1016, Page 272

activated scan chains in part2 (assume the length of the longest scan chain in two parts is the same).

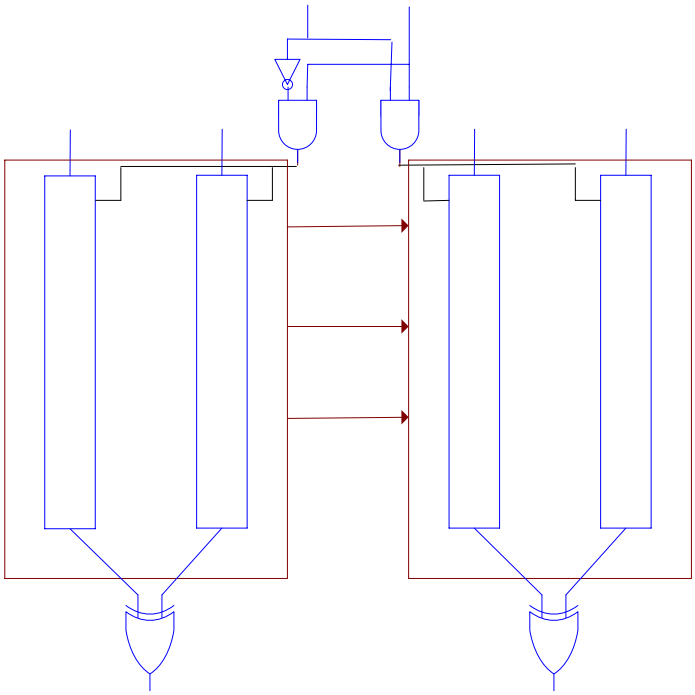


Figure 1: The Scan Architecture

Noting that keeping some scan chains active while others inactive has two meanings. Clock disabling allows only one part to be active, for example, part1. And scan-in disabling allows only a part of scan chains in part1 to receive test data, while other scan chains in part1 receive constant 0. These two methods together reduce the test power.

In Figure 1, two scan-ins of different parts can be connected to the same scan-in pin, because they scan in test data in different cycles. If the number of scan chains in one part exceeds the number of scan-in pins in ATE, the shift operation of one part should be cut into several steps. This can be easily realized by activating the clock signals of different scan chains driven by the same scan-in in different cycles, which will increase the control unit overhead. Capture operation is still completed in two sequential steps.

Here, we do not pay attention to the test response compaction, since the control of scan-outs is synchronous with that of scan-ins. And test response can be compacted by the traditional way, such as MISR, or connecting to the ATE directly. Let n1 and n2 be the number of scan-outs needed by part1 and part2, respectively. The number of scan-out pins required is max(n1,n2).

Partition into two parts

In this section, we partition scan flip-flops into two parts, and no flip-flop in the second part can reach any flip-flop in the first part. The detailed algorithm is described as below. Though the notion of partition is similar to the one in [4], the algorithm and objective are quite different. A new algorithm is introduced to partition the scan flip-flops efficiently.

In order to reflect the structural relationship of the scan flip-flops, a directed graph called s-graph is used to represent the dependent relationship. In the graph, every node stands for a scan flip-flop and every edge stands for the connection between them in the topology of the circuit.

Ex. 1016, Page 273

Definition: In the topology of a circuit, if there is a combinational path from the output of scan flip-flop S_i to the input of scan flip-flop S_j, it is called that S_i can reach S_j or S_j is reachable from S_i. This relationship is noted as S_i -> S_j.

The directed s-graph G (V, E) is used to describe the structural dependency of all scan flip-flops. In the s-graph, V stands for the set of scan flip-flops, and E stands for the directed edges. There exists a unidirectional edge from S_i to S_j in the s-graph if S_i -> S_j.

```
Divide_Circuit_To_Two_Parts()
Input: G(V, E), N which is the number of scan cells in the circuit, Q which is a queue.
Output: the two parts of scan cells P1, P2
1   for i ← 1 to N
2     for j ← 1 to N
3       if(Sj→Si)
4         put(Sj, Rlist[i]) ;
5   for i ← 1 to N
6   {
7     for j ← 1 to Nri
8     {
9       if(Rlist[i][j] is marked)
10        put Pre[j] into Pre[i] ;
11      else
12        put Rlist[i][j] into Q ;
13    }
14    while(Q is not empty)
15    {
16      Sp=pop(Q)
17      for every element St in Rlist[Sp]
18      if(St is not in Pre[i])
19        put St into Q and Pre[i] ;
20    }
21    mark Si ;
22  }
23  sort all Pre[i] by the size ;
24  find the max Pre[m] which is smaller than Ns/2 ;
25  for j← 1 to m
26    if |P1 ∪ Pre[j]|< N/2
27      Union(P1,Pre[j]) ;
28    else
29      Break ;
30  the scan cells which are not in P1 belongs to P2 ;
```

Figure 2: The algorithm to partition the scan flip-flops into two parts

The main idea of the algorithm is described as figure 2. The object of the algorithm is to partition scan flip-flops into two parts, then merge them to balance the two parts. First, for every scan flip-flop *i*, the set of flip-flops *Rlist* that reach *i* can be obtained. All sets *Rlist* can be obtained by traversing the circuit only once. Therefore, CPU time can be trivial. Then the breadth-first search algorithm is used to find all scan flip-flops that reach *i* directly or indirectly, as described from line 5 to 21. This can be implemented by

traversing the scan flip-flops which can reach it iteratively. Here a queue *Q* is used which is helpful in the BFS algorithm. The calculated set of each scan flip-flop is put into *Pre[i]*. After scan flip-flop *i* and all elements in *Pre[i]* are merged as *Pre[i]*, the scan flip-flops in *Pre[i]* constitute part1 and the remaining ones are put into part2. Finally, two smaller ones are merged to make sizes of two parts be balanced, as described from line 23 to 30.

Experimental result

The proposed scan testing method has been implemented. Experimental results on eleven larger ISCAS89 benchmark circuits are provided in Tables 1 and 2. In Table 1 and 2, test time, test data, test peak power, test average power and partition results are presented. Parameters FFs and NP represent the number of scan flip-flops, number of partitions, respectively. The parameters TA, AP, PP and CP represent the reduction ratios for test application time, average test power, and peak power, maximum capture power in a capture cycle, respectively. The parameters #P1 and #P2 represent the number of scan flip-flops in part1 and part2, respectively. The parameter #chains represents the total number of scan chains needed.

The ATPG tool ATALANTA is modified to generate test vectors for all circuits. Test power is estimated as the weighted number of transitions of all the nodes in the circuit. Here, the weight is the number of fan-outs to the node.

Table 1 lists the partition results for every circuit. From the result, it is shown that, the scan flip-flops can be clustered into two balancing parts by the proposed method.

Table 1: Partition result

circuit	FFs	#NP	#P1	#P11	#P12	#P2	#P21	#P22	#chains
s5378	179	2	42	42	0	137	137	0	9
s9234	228	3	113	56	57	115	115	0	8
s13207	669	3	325	146	179	344	89	255	10
s15850	597	4	279	134	145	318	318	0	10
s35932	1728	4	768	288	480	960	480	480	8
s38417	1636	4	816	326	490	820	360	460	10
s38584	1452	2	27	27	0	1425	1425	0	10

Table 2: Comparing with Full Scan

circuit	TA(r%)	AP(r%)	PP(r%)	CP(r%)
s5378	75.42	29.33	23.79	29.80
s9234	74.56	60.85	54.07	54.07
s13207	74.29	65.53	49.97	67.40
s15850	77.22	63.68	54.11	64.98
s35932	72.11	65.96	65.28	65.28
s38417	77.87	63.17	51.32	72.32
s38584	77.96	56.36	41.36	41.36
average	75.63			56.46

Table 2 presents the comparison between the proposed scan architecture and the single scan chain architecture. Power is calculated as the number of scan chain transitions. The peak power is usually determined by the maximum transitions in a single clock cycle. All circuits can be partitioned into two balancing parts except s38584. However, the peak

Ex. 1016, Page 275

power of s38584 is reduced greatly. The reason may be that filling strategy is useful to reduce the capture power.

In Table 3, the proposed new scan architecture is compared to the two previous methods [5] and [6]on test application time, average test power and peak test power. The test time is calculated as total clock cycles. Results are represented as reduction percentage. For the two methods, their results are given as the reduction ratio to their own standard full scan. From the table, it is shown that our method can reduce average power and peak power efficiently.

Table 3: Comparing with Others.

circuit	Method in [5] (r%)	Method in [6](r%)	The proposed method (r%)
s5378	58	*	29.80
s9234	50	57.61	54.07
s13207	59	55.34	67.40
s15850	34	49.09	64.98
s35932	23	40.06	65.28
s38417	40	61.48	72.32
s38584	53	56.01	41.36

Conclusion

This paper presents a low power architecture. All scan flip-flops are partitioned into two subsets, where test responses of one subset do not depend on the test data of the other subset. Experimental results show that the proposed test scheme can obtain better results on test power reduction than the previous methods.

References

1. M. L. Bushnell and V. D. Agrawal, Essential of Electronic Testing, Kluwer Academic Publisher,2000.
2. K. M. Butler, “Minimizing power consumption in scan testing: pattern generation and DFT techniques,” in Proc. IEEE Int. Test Conf., pp. 355-364, 2004
3. S. Wang and S. K. Gupta, “ATPG for heat dissipation minimization during test application,” IEEE Trans. Computers, vol. 47, pp.256-262, Feb. 1998
4. P. Rosinger, B. M. Al-Hashimi, and N. Nicolici, “Scan architecture with mutually exclusive scan segment activation for shift and capture power reduction,” IEEE Trans. on Computer-Aided Design, Vol. 23, no.7, pp.1142-1153, July, 2004
5. X. Wen, S. Kajihara, K.Miyase, T. Suzuki, K. K. Saluja, L. T. Wang, K. S. Abdel-Hafez, and K. Kinoshita, “A new ATPG method for efficient capture power reduction during scan testing,” Proc. of IEEE VLSI Test Symposium, pp 58-65,2006
6. Q. Xu, D. Hu, and D. Xiang, “Pattern-directed circuit virtual partitioning for test power reduction,” in Proc. of IEEE Int. Test Conference, Oct. 2007
7. X. Chen, Michael S. Hsiao, “An overlapping scan architecture for reducing both test time and test power by pipelining fault detection,” IEEE Trans. on VLSI System, vol. 15, no. 4, pp. 404-413, Apr. 2007.

ECS Transactions, 18 (1) 249-254 (2009)
10.1149/1.3096458 © The Electrochemical Society

A Self-Repairable Microprocessor

Yuanjiang Xie[a,b,c], Xiang Fu[a,b,c], Zijian He[a,b,c], Yang Zhao[d], Yu Hu[a,b], Xiaowei Li[a,b]

[a] Key Laboratory of Computer System and Architecture, Chinese Academy of Sciences, Beijing 100190, China
[b] Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[c] Graduate School of Chinese Academy of Sciences, Beijing 100049, China
[d] Department of Computer Science and Engineering, University of California, Riverside, 92521, USA

With increasing defect density, microprocessors, especially the embedded caches, will encounter more faults. Adding spare resources to replace defective components is a widely accepted method for yield enhancement. In this work, a repair method using content addressable memory combined with spare bits, as well as a novel fault injection method is proposed. With the proposed fault injection technique, various numbers and types of faults can be flexibly injected into the silicon. A wireless sensor network system using our self-repairable microprocessor (SRP) is developed to prove the effectiveness of the proposed technique.

INTRODUCTION

With advances of integrated circuit manufacture technology, geometry shrink and process variation tend to result in higher defect density. Meanwhile, another marked trend is embedded memories occupy more area in microprocessors and System-on-Chips (SOCs). According to the report of International Technology Roadmap for Semiconductors (ITRS) in 2000, the area ratio of memory will reach 94% of the embedded chip in 2014 [1]. Therefore, yield of memory becomes the key factor to impact chip yield. Prior research has indicated the yield of an 8Mbits memory without any repair methods is only 70% in the 0.13um standard technology node [2]. As build-in self-test (BIST) and build-in self-repair (BISR) of memory is a well developed and widely accepted concept for yield enhancement, there have been many efficient BISR methods proposed. Li and Huang proposed a high repair rate technique using spare rows and columns [3, 4]. Fu proposed to divide a spare column into multiple segments to increase repair rate [5]. Wu brought forward using spare words to repair faulty ones, the repair rate of which is very high [6]. And a method using spare word, spare rows and columns together is proposed in [7].

In this work, a fault injection technique and a self-repair technique based on content addressable memory (CAM) are proposed. Various numbers and types of faults can be flexibly injected into the memory by software instructions or hardware switches. Afterwards, BIST and BISR are activated to repair the faulty memory words. We also developed a wireless sensor network system using our self-repairable microprocessor (SRP) to demo the entire fault injection, BIST and BISR procedure.

The rest of the paper is organized as follows: Section 2 describes the fault injection technique, Section 3 briefly presents the redundant resources organization, and then the content addressable memory based BISR technique is proposed in Section 4. Section 5

Ex. 1016, Page 277

gives a demo of a wireless sensor network system using our self-repairable microprocessor, and finally Section 6 concludes the paper.

FAULT INJECTION

Since occurrence of manufacturing defects in an embedded memory usually presents with certain probability distribution, for example, the negative binomial distribution, a deterministic fault injection method is needed to validate the effectiveness of the proposed BISR technique. Conventional fault injection technique can be broadly classified into two categories, the one is tester-based techniques, a typical example is Focused Ion Beam (FIB) technique which use ion milling to cut defined metal lines or use deposition to create lines; the other one is simulation-based techniques such as fault simulation. Although tester-based techniques can inject defects into silicon chips, they are usually very time-consuming and costly. On the contrary, simulation-based techniques are easy to implement but are only applicable to circuits described in algorithms or hardware description languages. In this work, a fault injection technique is proposed to flexibly inject faults into the microprocessor chip with low cost.

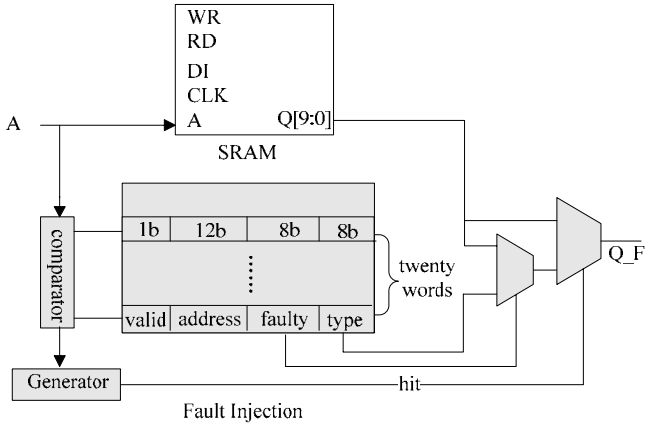


Figure 1 Structure of the Fault Injection Module

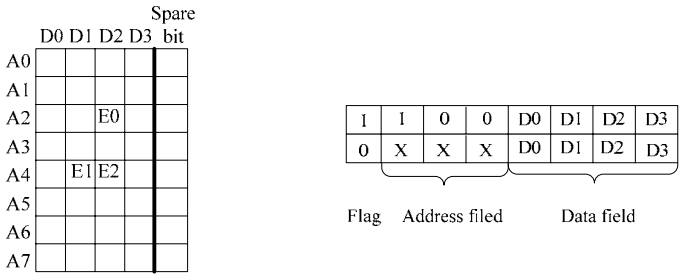
Figure 1 illustrates the structure of the fault injection module. The fault injection module is basically a CAM with twenty storage words, which means we can simulate at most twenty faulty words. Since the size of the embedded static random access memory (SRAM) in our microprocessor is 4096*8 bits, and we need to inject stuck-at-0 (SA0) fault and stuck-at-1 (SA1) fault into the SRAM, the word size of the fault injection CAM is 29 bits. As shown in Figure 1, the CAM word consists of one valid bit to indicate whether the CAM word is valid or not, twelve address bits to present the address of the SRAM word to be injected faults, eight fault flag bits to indicate whether the corresponding bit has been injected fault or not, and eight fault type bits to indicate which kind of faults the defective SRAM bit has. A '0' fault type bit means the faulty SRAM bit has a SA0 fault, while '1' indicating a SA1 fault. When accessing the SRAM memory, the accessed address is simultaneously compared with the address stored in the fault injection CAM. If there is a hit, then the data read out from SRAM will be converted based on the content of the CAM words. For example, if the content of the CAM word is 1,000011110000,00001001,00000001, it means we inject a SA1 fault to the bit 0 and a SA0 fault to the bit 3 of the SRAM word whose address is 000011110000. We can modify the content of the fault injection CAM by software instructions or hardware switches.

Ex. 1016, Page 278

REDUNDANT RESOURCES

Since the memory library provided by the foundry does not offer memories with spare rows or spare columns, we synthesized a spare bit for each word to repair single faulty bit in a word, as shown in Figured 2(a). It is an example of a SRAM with 8*4 bits and one spare bit. Assume there are three faulty bits in the SRAM, which are E0, E1, and E2. Except of the spare bit in each SRAM word, a CAM is also served as the redundant resource. Figure 2(b) shows the architecture of the CAM used to repair faulty words. The CAM also has a flag bit to indicate whether the CAM word is valid or not. The address field stores the address of the faulty SRAM word. The data field is used as a spare word to repair the faulty SRAM word if its faulty bits are more than one. We can see E0 is repaired by the spare bit, E1 and E2 are repaired by spare words provided by the CAM.

In our design, the original SRAM is 4096*8 bits. After redundant resources designed, the SRAM becomes 4096*9 bits. And the number of spare words provided by the repair CAM is ten.



(a) (b)

Figure 2 An example of simple SRAM ((a) An example of 8*4 bits SRAM with spare bits, (b) Architecture of CAM used to repair faulty words)

CAM-BASED BISR

In our work, a CAM-based BISR technique is proposed. The BISR architecture is shown in Figure 3. It contains two major parts: a bit repair unit to store spare bit shift information and a word repair unit to store spare word repair information.

The bit repair unit is composed of a CAM, a comparator and a hit signal generator. Each CAM word contains one bit of valid flag, twelve bits of SRAM faulty word address and eight bits of shift information. The valid bit indicates whether the CAM word is valid or not. In repair mode, if valid bit is 0, the CAM word can be used to store new repair information. In function mode, if the valid bit is 0, it means the CAM word has no repair information; otherwise, the CAM word stores repair information. The comparator will compare the accessed address with all valid addresses stored in the CAM word. And the generator will generate signal ‘hit_bit’ based on the comparison result. If there is a hit, signal ‘hit_bit’ will be asserted, and the data of Q will be shifted based on shift information stored in the CAM word. The shift operation skips the faulty bit in the data of Q and packs the other eight bits to Q_b. Notice the data written into the SRAM will be shift firstly to avoid writing data in the faulty bit if ‘hit_bit’ is valid.

Like the bit repair unit, the word repair unit is also composed of a CAM, a comparator and a hit signal generator. The function of the comparator and the generator is the same as those in the bit repair unit. The major difference between the word repair unit and the bit repair unit is the CAM of the word repair unit has a data field used as a

Ex. 1016, Page 279

spare word, while the bit repair unit has a shift field used to shift data. When the 'hit_word' signal is asserted, the data will be written in or read out from the data field of the CAM rather than the SRAM. The data field has eight bits.

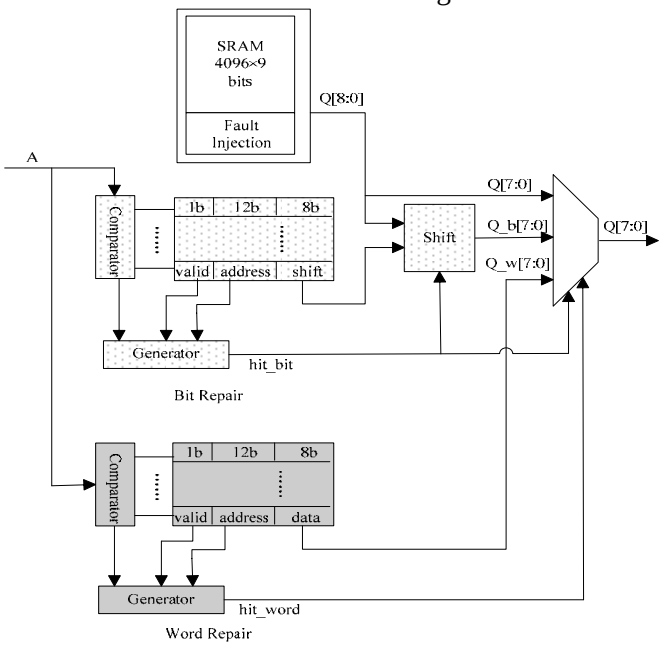


Figure 3 BISR architecture

The microprocessor is synthesized with Synopsys Design Compiler and the SRAM is synthesized with Artisan memory compiler using a 0.18 standard library. The area overhead of the microprocessor, as well as its SRAM, fault injection module, spare bits, test and repair circuitry (including BIST, BIRA and BISR) is 1662725.75 um^2, 311216 um^2, 74075 um^2, 38902 um^2 and 105928 um^2, respectively. The SRAM in the self-repairable microprocessor is placed as 256 rows, and every row has 16 words. All faults simulated are single cell faults. And the distribution of the injected faults is generated with Poisson distribution. As shown in Table I. Column "Spare Bits" and "Spare Words" indicates the number of CAM words of the bit repair unit and the word repair unit respectively that is needed to get 100% repair ratio.

TABLE I. Redundant Resources Used by the Proposed Method to Achieve 100% Repair Ratio.

Number of Injected Faults	Spare Bits	Spare Words
10	10	4
20	16	7
30	24	10
40	28	13
50	31	17

APPLICATION SCENARIO

We also developed a wireless sensor network system using our SRP chip to demo the entire fault injection, BIST and BISR procedure. Wireless sensor nodes can monitor the light intensity of the environment and transmit the data to a sink node which is responsible for transmitting data to a computer. Figure 4 shows the working flow of the microprocessor. The BIST module implements a word-oriented March17N algorithm ⇳{w0},⇧{r0,w1,r1,w0,r0,w1},⇧{r1,w0,w1}, ⇩{r1,w0,w1,w0}, ⇩{r0,w1, w0}. The

data background is '000000000'. Since the redundant resources are spare bits and spare words, we can use a simple redundancy analysis scheme in BIRA module: if the number of faulty bits in a SRAM word is only one, the faulty word is repaired by the spare bit, otherwise, it is repaired by a spare word. As shown in the figure, function of Fault Injection, BIST, BISR, BIRA modules can be controlled by a computer through transmitting data into the UART controller to generate control signals. Also these control signals can be controlled by hardware switches on the board.

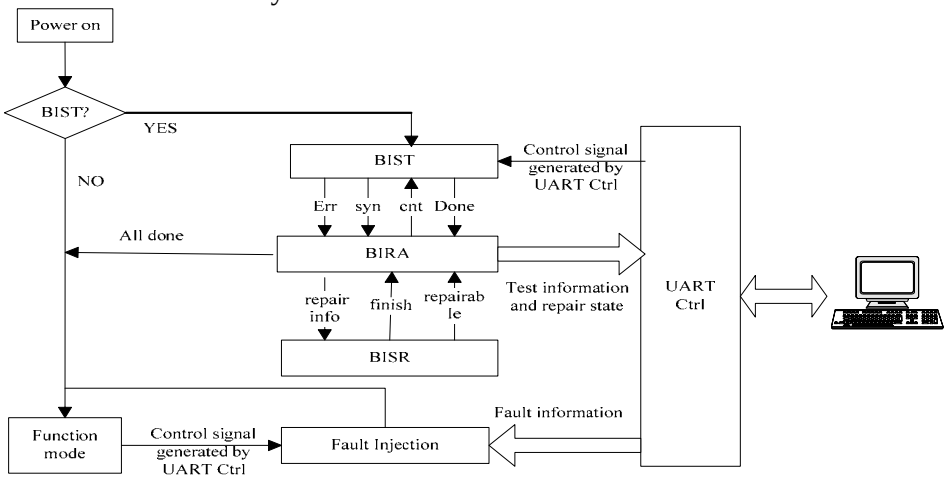


Figure 4 Working flow of the self-repairable microprocessor

A software tool named SNAMP is used to control data transmission and also send instructions to control the test and repair circuitry of SRP. Now, we can inject faults by SNAMP or by switches on the board. However, not all faults would affect the running of the application. Only faults occurred in certain addresses of SRAM will result in abnormal light intensity data that is displayed in the computer. Figure 5 is a snapshot of the running of SNAMP. The example system has three nodes, including a sink node, a SRP node indexed as node 2 and a normal node indexed as node 3. Since node 2 and node 3 are located in the same environment, their light intensity data should be close. In fact, as shown in Figure 5, the curve of node 3 and the curve of node 2 are very close. After injecting some faults into SRAM, the curve of node 2 falls down significantly. However, it will recover to normal mode after BISR.

CONCLUSION

A self-repairable microprocessor is developed in our work. The novelty of the design contains two aspects. One is the design of a fault injection module which is used to simulate memory faults. And the other one is the CAM-based self-repair technique which does not depend on spare resources provided by the manufacture. We develop an embedded system to demo how seriously the faults in the microprocessor memory affect the running of the application and the embedded system. By using the proposed BIST and BISR technique, the microprocessor runs well after repair as long as it is repairable.

Acknowledgments

This work is supported in part by National Natural Science Foundation of China (NSFC) under grant No. 90707004, 90607010 and 60633060, and in part by National Hi-

Tech Research and Development Program of China (863 Program) under grant No. 2007AA01Z107, 2007AA01Z113.The authors would like to thank Da Wang, Hongsong Zhu, Yongjun Xu, Chaonong Xu, Shirong Lu, Ze Zhao for their valuable comments, constructive suggestions and help of the circuit design and the wireless sensor network system design. This work was done when Yang Zhao was at Institute of Computing Technology, Chinese Academy of Sciences.

Corresponding authors: Yu Hu and Xiaowei Li, {huyu, lxw}@ict.ac.cn

References

1. International Technology Roadmap for Semiconductors. DOI=http://public.itrs.net/Files/2000UpdateFinal/2kU dFinal.htm
2. Y. Zorian and S. Shoukourian, Embedded-Memory Test and Repair: Infrastructure IP for SoC Yield. *IEEE Design & Test of Computers*, p. 58, 20, 3(2003)
3. J.-F. Li, J.-C. Yeh, R.-F. Huang, et al., A built-in self-repair design for RAMs with 2-D redundancy. *IEEE Transactions on VLSI Systems*, p. 742, 13, 6(2005)
4. C.-T Huang, C.-F. Wu, J.-F. Li, et al., Built-in redundancy analysis form memory yield improvement. *IEEE Transactions on Reliability*, p. 386, 52, 4 (2003)
5. X. Fu, D. Wang, H. Li, et al., A built-in self-repair mechanism for embedded SRAM. *In Proceedings of China Test Conference (in Chinese), Beidaihe*, p. 15 (2006)
6. Z. Wu, X. Zhou, M. Lei, et al., A built-in self-repair design for embedded memory. *Journal of Microelectronics & Computer (in Chinese)*, p. 79, 24, 2 (2007)
7. S.-K. Lu, C.-L. Yang, H.-W. Lin, Efficient BISR techniques for word-oriented embedded memories with hierarchical redundancy. *In Proceedings of IEEE/ACIS International Conference on Computer and Information Science and IEEE/ACIS International Workshop on Component-Based Software Engineering, Software Architecture and Reuse*, p. 355, Honolulu (2006)

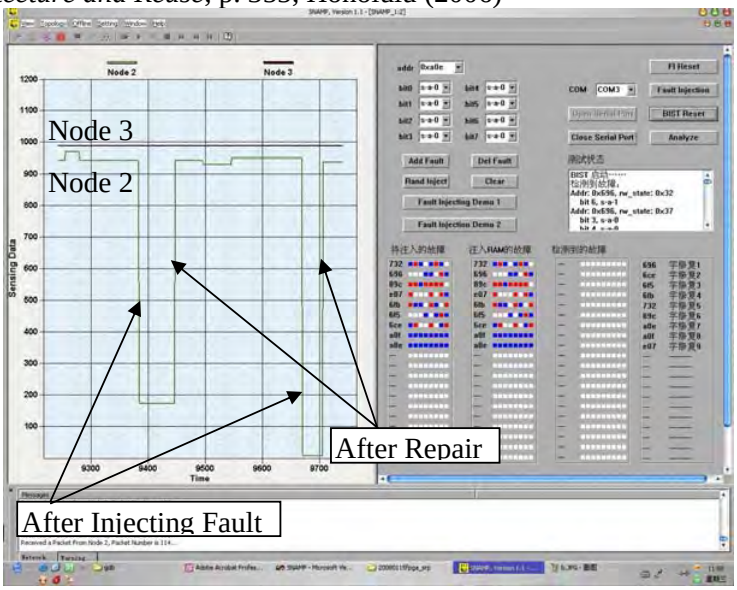


Figure 5 A snapshot of the running of SNAMP

ECS Transactions, 18 (1) 255-260 (2009)
10.1149/1.3096459 © The Electrochemical Society

Improved Cost of Test by Optimized Tester Utilization

Günther Jeserer

Multitest elektronische Systeme GmbH, Rosenheim, Germany

The semiconductor test industry agrees that cost of test (CoT) is one of the biggest challenges today and it will be crucial in future. Test cell efficiency needs to be increased. Enhanced tester capabilities allow shorter test times and higher parallelism. Handlers try to keep pace with higher parallelism but missmatch with speed. The urgent challenge is to optimize the tester utilization in order to benefit from higher parallelism as well as from shorter test times. The solution needs to improve the ***overall*** handler preformance. Increasing the number of contact sites will not be sufficient. In addition speed related parameters like soak capacity, index time and loading and sorting speed need to be considered.

Cost of Test

Market forecasts[1] state that the average sales price per IC will continue to drop. Thus production cost need to be reviewed. As for many applications one cannot give up final test, there is a clear industry-wide target to minimize cost of test (CoT)[2].

Cost of Test Model

This paper will refer to a comprehensive CoT model dedicated to final test. It is based on common management ratios e.g. "Yield", "Utilization", "Capital Cost" etc. but then broken down into test cell specific figures. For simplicity in this paper "test cell" is used as a synonym for tester plus test handler and "IC", "Package" or "DUT" for the device to be tested.

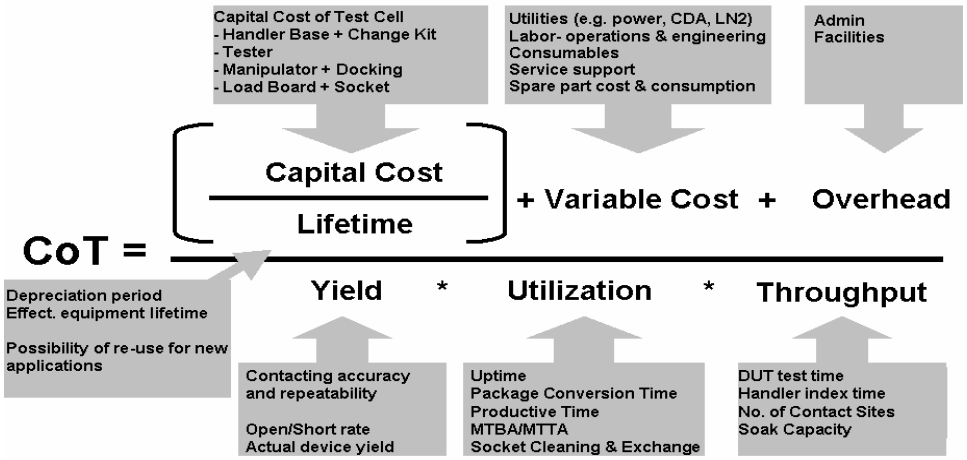


Figure 1. Comprehensive cost of test model

Ex. 1016, Page 283

Test cell throughput as key to improved CoT

Although changing any of these CoT parameters will have an impact on overall the CoT, here the focus will be on throughput and its determinants. With reference to the introduced CoT model "*reducing* CoT" requires to *increase* the throughput of the test cell. Whereas the testers already offer enhanced capabilities, the DUT handlers currently seem to be the bottle neck for optimizing the test cell performance. Therefore the following consideration will only deal with throughput relevant determinants of the test handler: DUT test time, handler index time, number of contact sites, soak capacity.

Throughput is stated in unit per hours (uph).

Test Cell Processes and Interdependencies

Looking generally at the overall test cell, "final test" is a chain of alternating sequences "test" and "DUT exchange". (Figure 2.)

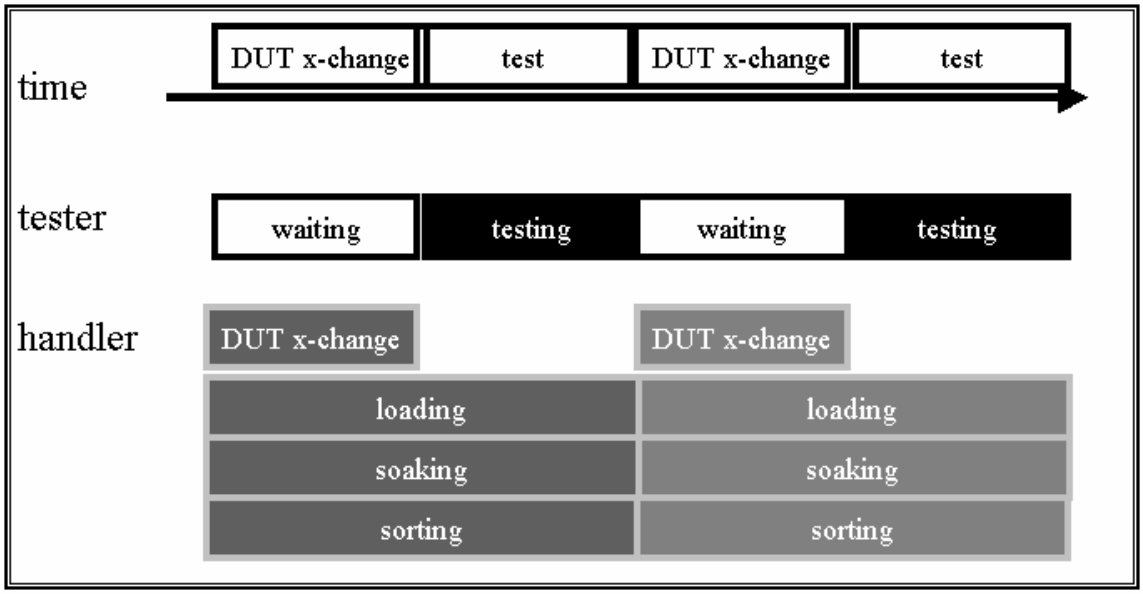


Figure 2. Schematic process sequences of the test cell

Tester

As we take the tester performance as given, it is sufficient just to differentiate between times, where the tester is busy and times where it is idle – waiting for the handler.

DUT handler

In order to analyze the handler limitations the process steps of the test handlers are broken down into:

- Loading
- Soaking
- DUT exchange
- Sorting

Terms

For the subsequent analysis it is important to define the following key terms (Figure 3.)

- Test time: Time the tester needs to perform the test
- Index time: Time the handler needs to replace the tested IC by a not yet tested IC at the contact site
- SOT: Start of test, start of the tester's activity
- EOT: End of test, end of the tester's activity

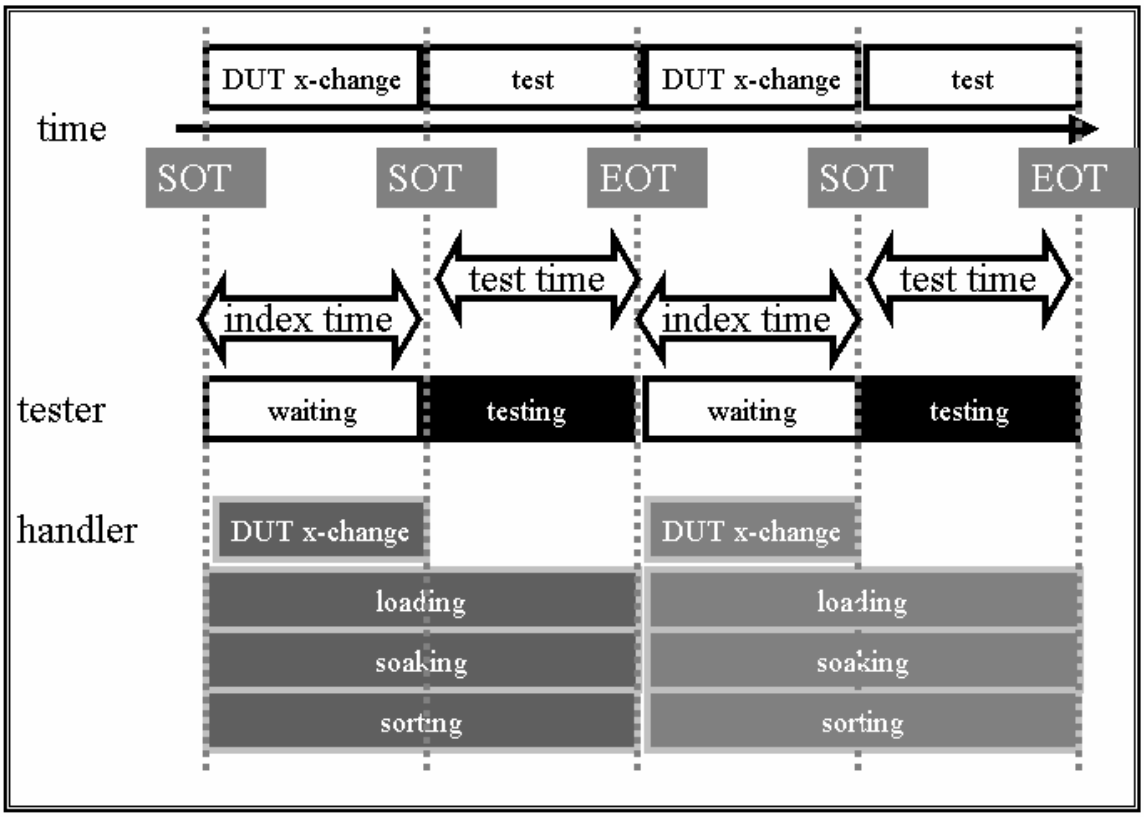


Figure 3. Graphic description of key terms: "test time", "index tme", "SOT" (start of test), "EOT" (end of test)

First conclusions:

In today's state-of-the-art test cells the tester is already waiting for the handler during the index time i.e. the tester is idle, non productive. This issue gets even worse with improved tester capabilities like shorter test times (Figure 4.) and higher parallelism (Figure 5.). Shorter test times give the handler less time for loading, soaking and sorting, whereas higher parallelism requires to load, soak and sort more ICs for one test cycle. Given unchanged handler capabilities this means more load, soak and sort cycles for one contacting cycle.

Ex. 1016, Page 285

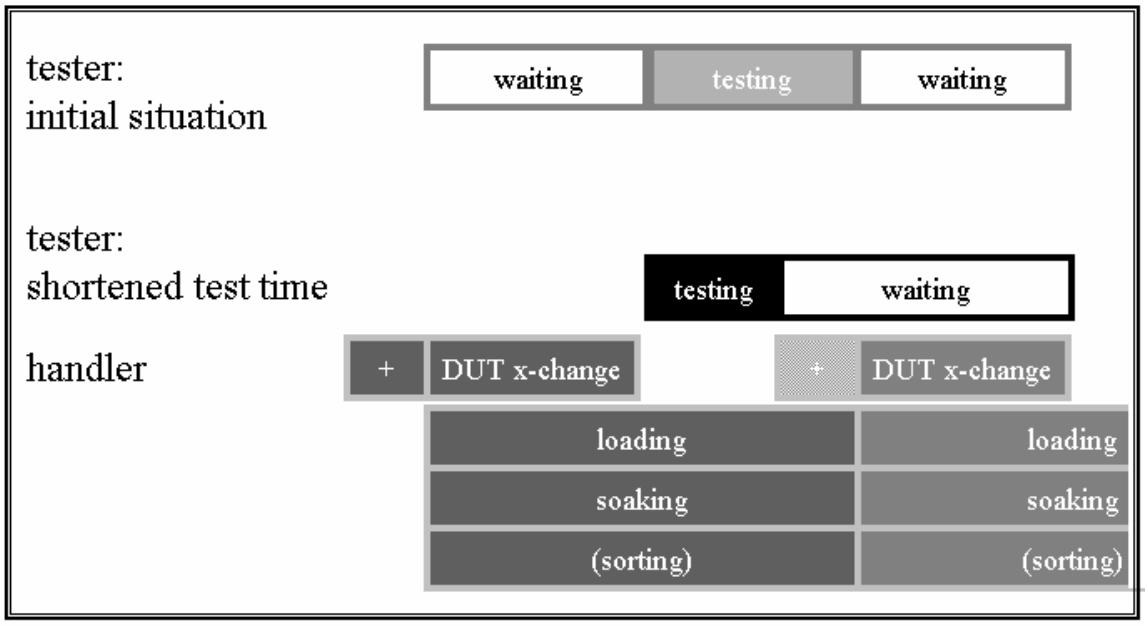


Figure 4. Impact of shorter test times

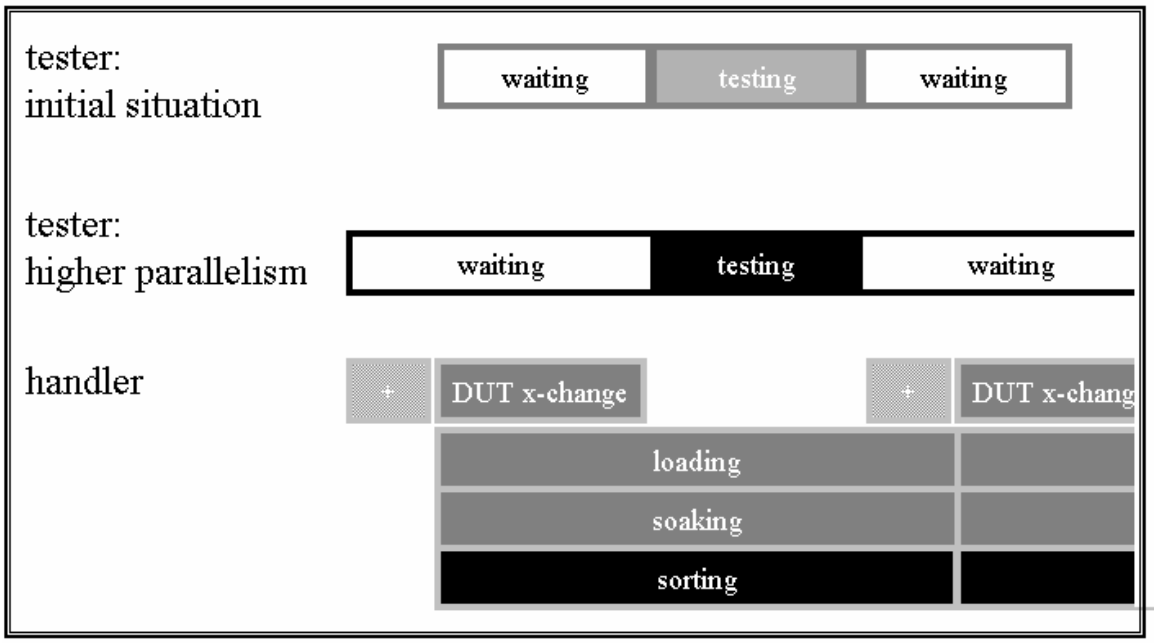


Figure 5. Impact of higher parallelism

Optimizing for Throughput

In the ideal situation with no limitations by the handler the enhanced tester capabilities allow to optimize for throughput in two ways (Figure 6):

1) benefit from shorter test times: uph(t_2) vs. uph(t_1)
2) use the higher parallelism: uph(p_2) vs. uph(t_1)

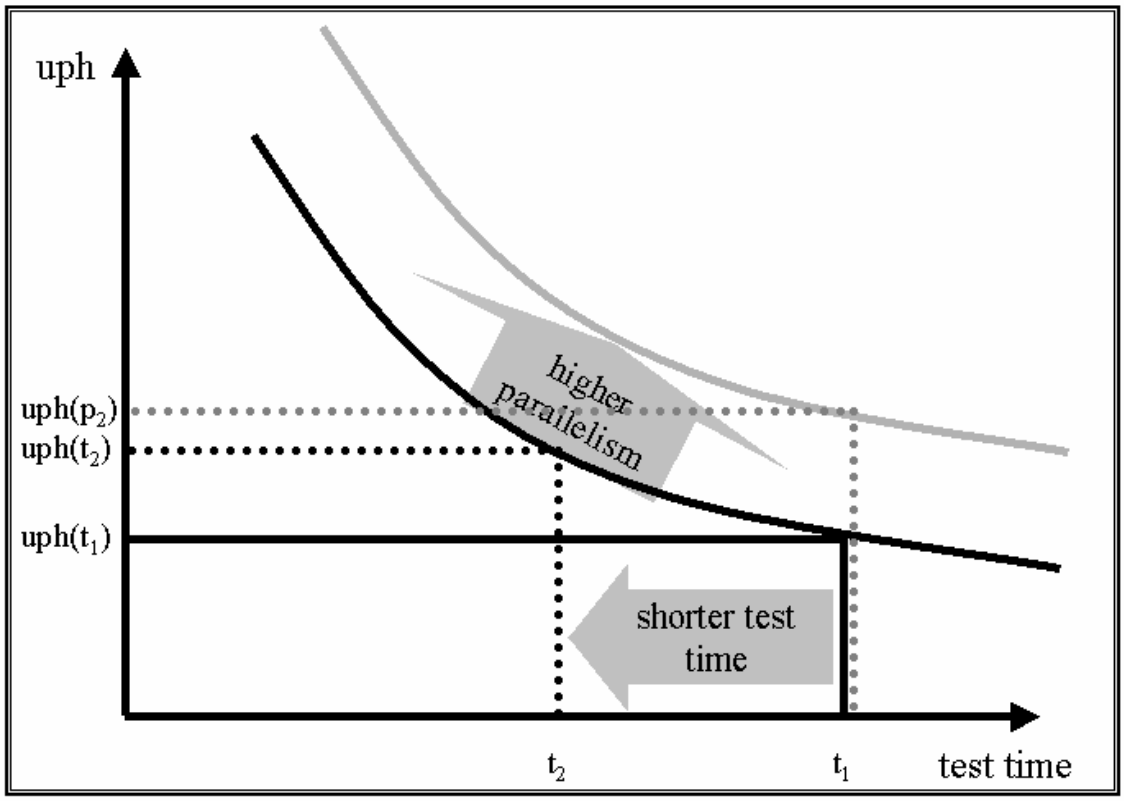


Figure 6. Optimizing for throughput: Test time and parallelism

Limitations by the handler

Unfortunately the current state-of-the-art handlers cannot follow the enhanced tester capabilities even if they offer a corresponding number of contact sites. Thus the possible throughput advantages are capped by the maximum handler throughput determined by the speed for loading, soaking and sorting. (Figure 7.)

Number of contact sites. Increasing the number of contact sites (= increasing parallelism) will directly respond to the enhanced tester capability of testing more devices simultaneously.

Handling speed/capacity: Increasing parallelism is only sufficient as long as the handler maximum capacity for loading, and sorting is not reached otherwise it will be offset by longer loading and sorting times. In any case the handling speed needs to be improved to benefit from shorter test times.

Soaking capacity: The same is true for the needed soaking capacity to fully leverage the higher parallelism and shorter test time.

Conclusion

In order to really leverage the full tester capabilities test handlers do not only need to be optimized for single parameters (parallelism, speed), but need to be synchronized

internally with respect to the number of contact sites, the loading, soaking and sorting capacity.

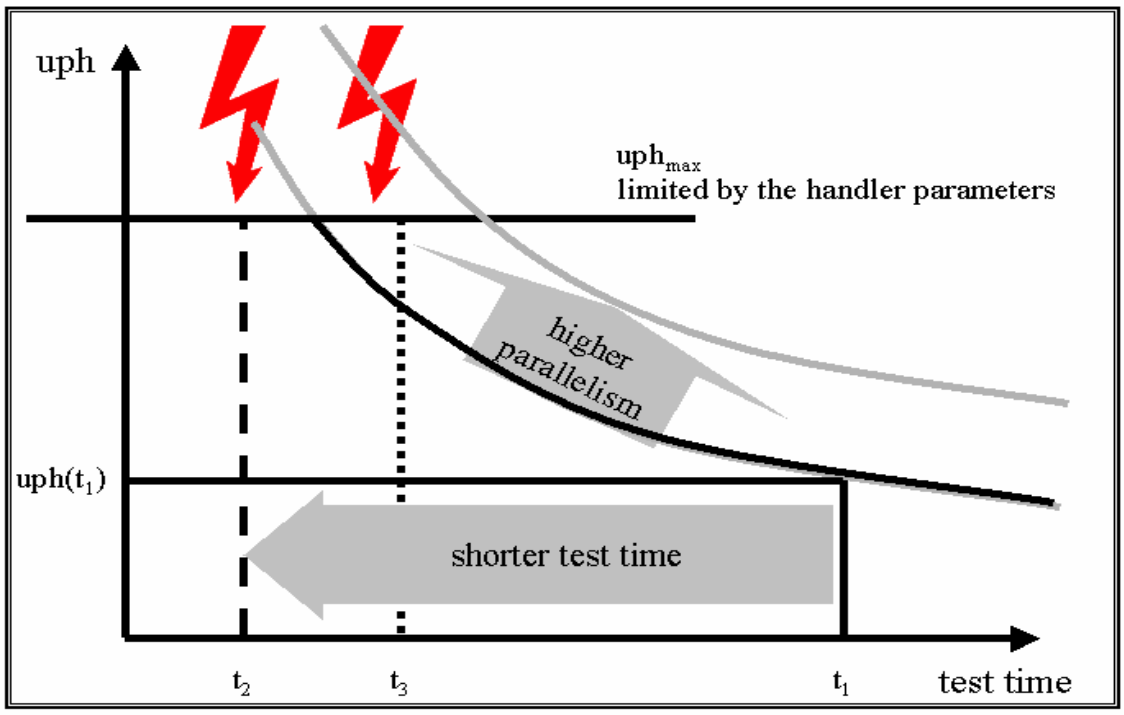


Figure 7. Limiting handler parameters

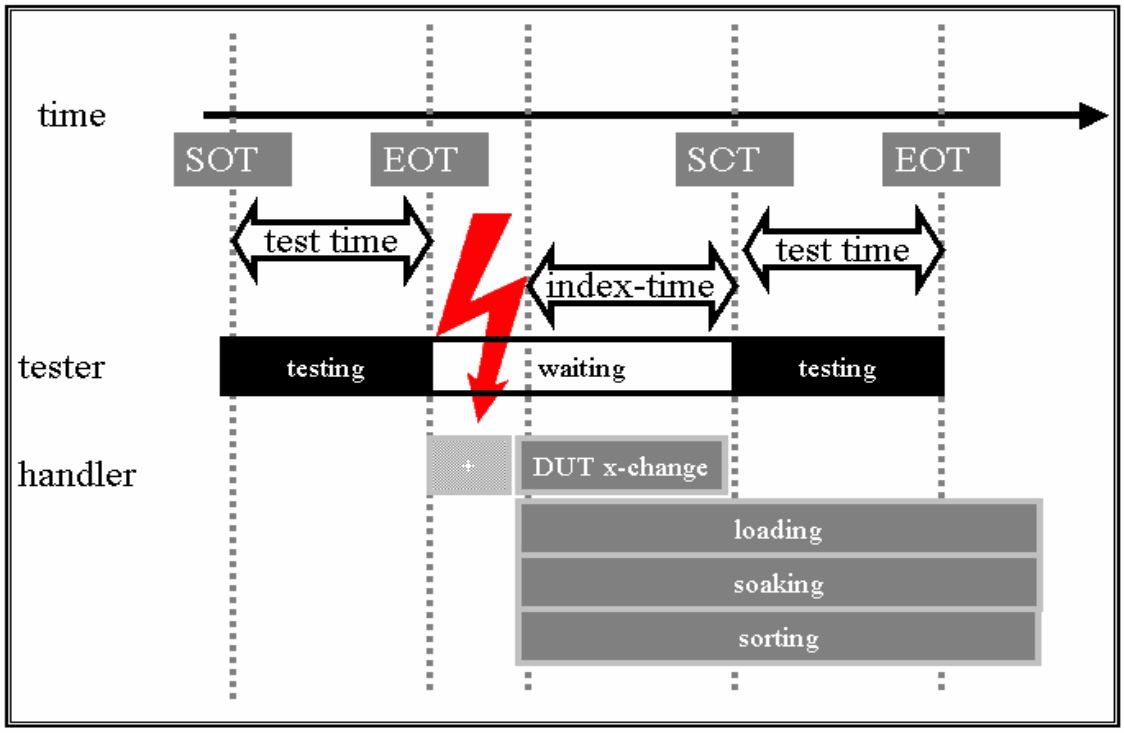


Figure 8. Limiting handler parameters

References

[1] VLSI Research; *www.vlsiresearch.com*

[2] Rudy Garcia, NPTest, *"Redefining Cost of Test In an SOC World"*, 2003, EE-Evaluation Engineering

CHAPTER 3

NEW MATERIALS AND PROCESS INTEGRATION

ECS Transactions, 18 (1) 263-268 (2009)
10.1149/1.3096460 © The Electrochemical Society

Semiconductor Cleaning Technology –Past and Future

Jerzy Ruzyllo

Department of Electrical Engineering and Department of Materials Science and Engineering
Penn State University, University Park, PA 16802, USA

This paper first briefly reviews selected aspects of the evolution semiconductor cleaning technology underwent over the years then considers challenges it is facing as the new material systems and innovative complex device configurations are being explored by semiconductor industry. Emerging trends are identified and methods developed to assure adequate performance of wafer cleans are outlined. The technique using mixture of anhydrous HF (AHF) and alcoholic solvent is considered as an example of a process which can offer solution to various future challenges.

Introduction

The wafer cleaning is the most frequently applied processing step in a typical semiconductor device manufacturing sequence. In the case of silicon, chemistry and implementation of cleaning operations are very well established and backed by many years of extensive research, as well as significant industrial tool base. As a result, silicon cleaning technology is by far the most mature among all semiconductors of any practical importance. The first complete, based on scientific considerations cleaning recipe specifically designed to clear Si surface from particles, metallic, and organic contaminants was proposed in 1970 [1]. Since then, silicon cleaning technology was undergoing continuous evolutionary modifications including early attempts to replace parts of the cleaning operations carried out in wet chemistries with gas-phase equivalents (e.g. [2-5]). Surprisingly, state-of-the-art Si cleaning still relies on roughly the same set of chemical solutions, but the way they are prepared and delivered to the wafer is very different from the originally proposed. In addition, selected surface cleaning/conditioning functions that were traditionally performed by wet cleaning chemistries are now carried in the gas-phase.

As advanced as they currently are, silicon cleaning methods can not meet all the diversified emerging needs of semiconductor technology across the spectrum of materials and device structures both in terms of implementation methods and chemistries. The ranges of the new materials being introduced into semiconductor device technology as well as variety of new non-planar device structures pose serious challenge to cleaning technology.

The purpose of the brief overview is to discuses past developments, current trends and future needs of semiconductor wafer cleaning technology. This review is continuation of the previously presented contributions summarizing needs and outlining challenges facing state-of-the-art semiconductor cleaning technology [6,7].

Ex. 1016, Page 291

Evolution of Semiconductor Cleaning Technology

As indicated earlier, first complete, based on scientific considerations Si surface cleaning recipe was introduced almost forty years ago [1]. Since then, the semiconductor cleaning methods evolved from the collection of mostly experimental techniques, based on a trial-and error approach, into a scientific and technical domain of critical importance to the manufacturing yield in particular and sustained growth of semiconductor industry in general. This trend is captured in the variety of review articles (e.g. [8]) as well as monographs (e.g.[9,10]) devoted to semiconductor cleaning technology. The detailed discussion of the evolution of semiconductor cleaning technology over the years would go beyond the scope of this paper. Hence, only selected key elements are outlined briefly below.

Objectives of Cleaning With time the focus of semiconductor cleaning technology was changing. In early years it was on gross particulate and metallic contamination, although, in reality the malfunctions of semiconductor devices were more often resulting from the high density of defects in the substrate wafer rather than from surface contamination. With particles and metallic contamination orders of magnitudes smaller than in early years, and contamination control in this regard by far superior, more attention is being paid now to organic contamination as well as to the issues related to broadly understood surface conditioning. The good news is that the efficient methods allowing control of organic contaminants are available. As Fig. 1 indicates with a simple lamp-cleaning [11,12] organic contaminants can be removed from the Si surface. In addition, the role ozone both dissolved in water and in the gas-phase plays in organic contamination control should be emphasized.

A separate issue is an observed diversification of the objectives of cleaning methods because of the different needs of FEOL and BEOL cleaning requirements with focus on post-CMP cleans in the latter case.

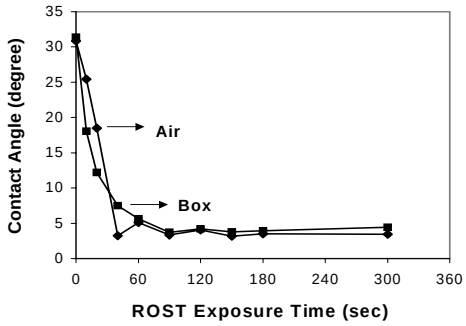


Fig. 1 Silicon surface contact angle changes as a function of lamp cleaning time.

Chemistries As far as cleaning media are concerned the wet cleaning remains a “work horse” of state-of-the-art wafer cleaning processes. While cleaning chemistries in Si technology did not depart far from the original RCA recipes [1], the use of heavily diluted solutions, focus on process simplification, and broad use of ozonated water are the most obvious modification of the overall process.

The APM (NH_4OH:H_2O_2:H_2O) based chemistry still dominates in particle removal, but without megasonic enhancement it would not be efficient enough. The metal removing

Ex. 1016, Page 292

chemistries based on the original RCA (HPM – $HCl:H_2O_2:H_2O$) recipe are all but abandoned. Innovative chemistries along with much cleaner resist and chemicals are the key to the success in this area. Also, the contamination control needs of very tight geometry device manufacturing are being served by innovative techniques involving for instance supercritical CO_2 (SCCO2) cleaning,

As far as dry cleaning is concerned it is being used in select, mostly surface conditioning steps. An example is an anhydrous HF (AHF)/alcoholic solvent process effective in removing native/chemical oxide from the Si surface in a variety of applications [13]. Figure 2 shows schematics of the 25-wafer and 3-wafer commercial reactor [14] implementing AFH/alcoholic solvent process in various surface processing operations. As AFM results in Fig. 3 show the process is implemented without damage to the Si surface.

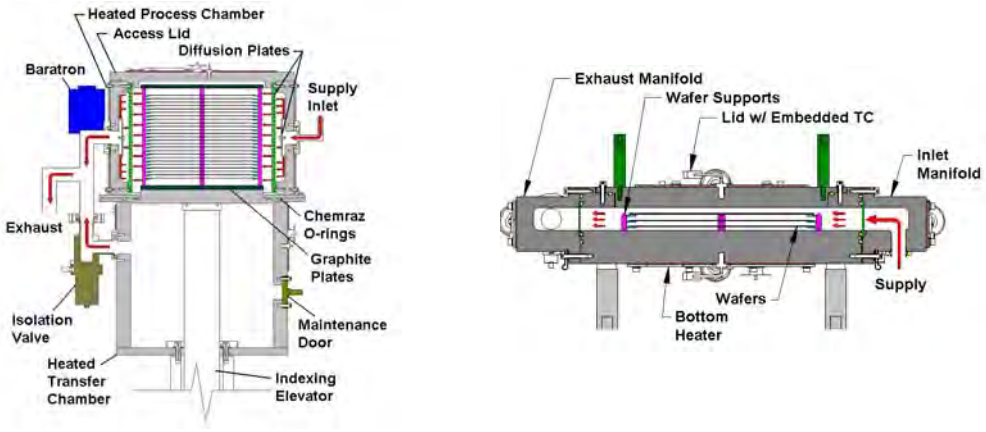


Fig. 2 Schematics of 25-wafer and 3-wafer AHF/alcoholic solvent process tool [14].

Wafer drying is an inherent part of any wet cleaning sequence. Significant progress in this key component of the cleaning process was accomplished by the broad acceptance of IPA-based wafer drying methods which employ Marangoni drying and its derivatives.

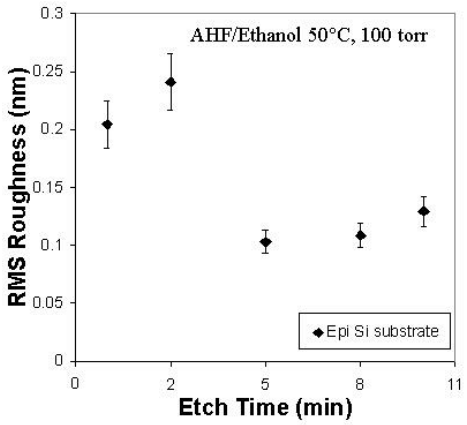


Fig. 3 Roughness of epi-Si surfaces vs. AHF/Ethanol etch time.

Tools and Methods The traditional immersion cleans continue to play dominant role both in batch as well as single-wafer cleaning processes. In the latter case the trend is

Ex. 1016, Page 293

reinforced by the needs of solar cell cleaning technology in which, due to the shear number of processed substrates, batch processing is an approach of choice. The reactor shown in Fig. 2 demonstrates that selected gas-phase cleaning chemistries can be made fully compatible with batch processes. At the same time, however, single-wafer cleaning methods, such as spin cleaning, are being pursued in the range of high-end applications.

Current and Future Challenges

In order for the semiconductor cleaning technology to meet emerging needs, various adjustments and modifications of current processes will have to be implemented [6.7]. To exemplify the nature of the potential problems the discussion below considers three among several emerging challenges.

Cleaning Related Damage As vertical dimensions continue to shrink, the material losses and surface roughening during cleaning operations is emerging as an area of key concern in surface preparation technology. The process of particle removal without material loss and pattern damage must involve serious tradeoffs. Physical aids such as megasonic agitation, potentially a key problem in terms of structural damage, patter collapse, etc., is being re-engineered so that efficiency of the particle removal process will be maintained without any adverse impact on pattern integrity. Considering the fact that an atomic scale deterioration of surface morphology may have a "killer" effect on device performance, even seemingly the most benign elements of the cleaning sequence, such as DI water rinse, must be re-evaluated from this point of view.

To curtail pattern collapse and related damage in some device structures vapor phase chemistries, such as discussed earlier anhydrous HF (AHF) mixed with the vapor of an organic solvent, are expected to become increasingly useful.

Non-Planar Devices As stated earlier, in order to respond to the issue of non-planarity of silicon surfaces, wafer cleaning technology is being challenged on at least three different fronts.

First is concerned with CMOS processing. At the continuously reduced device geometry the challenge in cutting edge digital CMOS technology is to maintain adequate capacitance density of gate structure needed to sustain sufficiently high drive current at the reduced gate length. One approach is to use gate dielectrics featuring higher than SiO_2 dielectric constant, the other is to increase gate area without increasing area of the cell by structuring MOS gate 3-dimensionally, and yet the other, the combination of both.

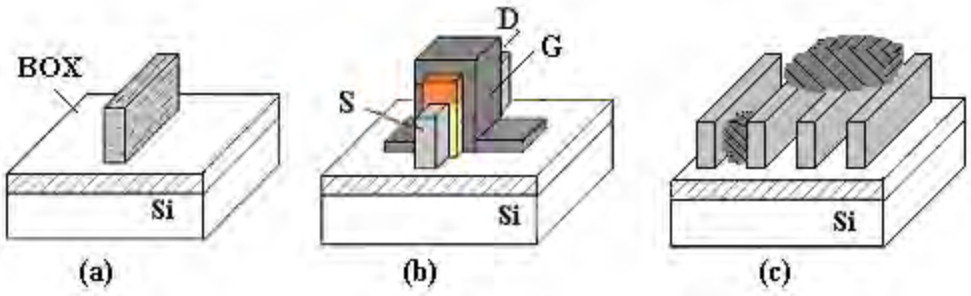


Fig. 4. (a) A "fin" defined in Si active layer of an SOI wafer, (b) MOSFET built around such "fin", (c) particles on the surface of a multi-fin structure.

Regardless of which specific solution will become a standard, the pre-gate oxidation cleaning and conditioning of post-RIE densely spaced vertical “fins” engraved in SOI silicon (Fig. 4a) will become an important element of the process. Figure 4b shows an example of an MOSFET built around a “fin”.

Different set of challenges is posed by MEMS (Micro Electro-Mechanical Systems) processing. Specific feature of MEMS manufacturing is that it includes deep etching of elaborate 3D features, as well as demanding release processes in which buried oxide is deep-etched laterally. Removal of possible etch residues from such extremely confined geometrical features and assurance of stiction-free operation of cantilever beams and membranes can not be accomplished using conventional wet etching and cleaning technology. The anhydrous HF/methanol (AHF/MeOH) (Fig. 2) sacrificial oxide etch process has been investigated as a viable solution to the latter [15].

Still different in nature problems are encountered when processing nano-geometry structures that depart from conventional planar technology. Making functional devices out of nanowires requires subjecting them to the sequence of operations very similar to those used in planar device processing. Considering extreme fragility of nanowires and a size that makes their handling very difficult, the use of the established fabrication techniques in a conventional fashion is restricted. In particular, the restrictions apply to wet cleaning/native oxide etching operations which, due to the problems of nanowires handling in liquids, are incompatible with nanowire processing.

Beyond Silicon Growing needs for improved performance in specific electronic (e.g. high-temperature, high-power, as well ultra-high speed) and photonic (e.g. emission of blue light or UV detection) applications, require significantly improved manufacturing technology in the range of semiconductors other than silicon. Examples of such materials include germanium, Ge, due to its higher than in Si electron mobility and possible integrateability with high-k gate dielectrics, silicon germanium, SiGe, needed to process strained-channel Si MOSFETs, as well as silicon carbide, SiC, for its wide energy gap. Also of growing interest are III-V semiconductors beyond the most advanced GaAs such as GaN, InAs, InSb, ZnO and several others.

Surface cleaning is becoming a growing issue in the processing of this type of semiconductors. This is because the inferior quality of the substrate crystal rather than cleanliness of its surface is no longer a dominant factor defining manufacturing yield with those materials. With the improvements in the quality of single-crystal of various semiconductor substrates the paradigm is shifting and much closer attention to the cleaning technology is being paid.

Semiconductor device technology rapidly expands into areas well outside of the mainstream silicon logic and analog applications. In display technology, solar panel technology and several other large-area electronic and photonic systems, the materials of which surfaces need to be processed may include glass, ITO (indium tin oxide), or flexible plastic substrates just to mention a few. Even in the mainstream silicon IC and III-V photonic applications the non-semiconductor substrates offer significant benefits, and hence, are vigorously pursued. For instance, the sapphire (single crystal Al_2O_3) is a substrate of growing importance in semiconductor device manufacturing. Cleaning of all these materials is an issue of significant interest

Ex. 1016, Page 295

Summary

The silicon cleaning technology developed over the years is a foundation upon which solutions to the challenges to surface processing of other semiconductors are based. The emergence of new materials of highly diverse chemical composition and material systems departing from the planarity and rigidity of standard wafers will enforce development of new approaches to semiconductor cleaning technology.

Acknowledgments

Thanks are due to Primaxx, Inc., and Sumitomo Precision Products Co., Ltd. for providing AHF tool used to obtained results presented in this paper and for an overall support of the dry cleaning projects.

References

1. W. Kern and D. Puotinen, *RCA Review*, 31, 87 (1970)
2. J. Ruzyllo, G. Duranko, and A. Hoff, J. Electrochem. Soc., 134, 2052 (1987).
3. J. Ruzyllo, *Microcontamination*, 60(3), 39 (1988).
4. J. Ruzyllo, D. Frystak, and R. Bowling, *1990 IEEE International Electron Device Meeting*, San Francisco, CA, Technical Digest, p. 409., 1990.
5. Y. Ma, M.L. Green, K. Torek, J. Ruzyllo, R. Opila, K. Konstadinidis, D. Siconolfi, and D. Bransen, , *J. Electrochem Soc.*, 142, L217 (1995)
6. J. Ruzyllo, in *Microelectronics Technology and Devices – SBMICRO 2007*, J.A. Martino, M.A. Pavanello, and C. Claeys Editors, ECS Trans, **9**(1), 87 (2007)
7. J. Ruzyllo, R.E. Novak, T. Hattori, P. Mertens. And P. Besson, in *Cleaning. and Surf. Cond. Technol. In Semicon. Dev. Manufacturing X* , T. Hattori, R. Novak, J. Ruzyllo Editors, ECS Transactions, 11(2), 3 (2007).
8. M. Heyns, P. Mertens, J. Ruzyllo, and M.Y.M. Lee, "Advanced Wet and Dry Cleaning Coming Together for Next Generation", *Solid St. Technol*., 42 (3), 37 (1999)
9. W. Kern (Ed.), Handbook on Semiconductor Cleaning Technology, Noyes Publications, 1993.
10. T. Hattori, (Ed.), Ultraclean Processing of Silicon Wafers, Springer Werlag, Heidelberg, 1998.
11. A. Danel, C.-L. Tsai, K. Shanmugasundaram, F. Tardif, E. Kamieniecki and J. Ruzyllo, "Cleaning of Si Surfaces by Lamp Illumination", *Solid State Phenomena*, Vol. 92, 195 (2003)
12. C-L.Tsai, P. Roman, C.-T. Wu, C. Pantano, J. Berry, E. Kamieniecki and J. Ruzyllo, "Control of Organic Contamination on Silicon Surface Using White Light Illumination", *J. Electrochem. Soc.*, 150, G39 (2003).
13. P. Roman, K. Torek, K. Shanmugasundaram, P. Mumbauer, D. Vestyck, P. Hammond, and J. Ruzyllo, *Solid State Phenomena*, vol.145-146 (2009)
14. Technical Literature, Primaxx, Inc., Allentown, Pennsylvania.
15. M.Erdamar, K. Shanmugasundaram, P. Roman, K.Chang, M. Klimkiewicz, and J. Ruzyllo, , *J. Micro- and Nanolithogrpahy, MEMS and MOEMS*, 7(3), 1 (Jul.-Sep. 2008).

ECS Transactions, 18 (1) 269-274 (2009)
10.1149/1.3096461 © The Electrochemical Society

Electromigration Improvement for Advanced Technology Nodes

Hui-Jung Wu, Wen Wu, Roey Shaviv, Mandy Sriram, Anshu Pradhan, Kie Jin Park, Jennifer O'Loughlin, Kaushik Chattopadhyay, Tom Mountsier and Girish Dixit

Novellus Systems, Inc., San Jose, CA, USA

Electromigration (EM) reliability has become a growing concern for Cu interconnects in advanced technology nodes. The interface between the dielectric diffusion barrier (SiCN) and copper, metal barrier quality and metallization homogeneity are all critical to EM performance. Our EM enhancement investigation has focused on improving the SiCN/Cu interface, metal barrier quality and metallization homogeneity while mitigating the process integration risk and cost. Multiple processes developed at Novellus have demonstrated significant EM improvement without compromising stress migration or dielectric reliability. The results of these process developments and the mechanisms of EM improvement will be discussed in this paper.

Introduction

Cu interconnect is now widely used in both logic and memory devices. With interconnect dimensions rapidly shrinking, following Moore's law, meeting reliability requirements becomes more challenging. Numerous approaches, ranging from evolutional process optimization to novel process/materials adoption, are being explored to improve EM lifetime. Among these, dielectric diffusion barrier (SiCN) and metal barrier/seed processes have been the main focus. Enhancing the SiCN/Cu interface through plasma pretreatment and optimizing the key SiCN film properties have been widely adopted (1). Metal barrier process optimization focusing on barrier step coverage, with re-sputtering, anchoring depth and via clean, have been reported previously (2,3,4). Innovative approaches have also been proposed to improve EM through the adoption of novel materials/processes. Selective metal cap (CoWP) improves EM through the replacement of the critical SiCN/Cu interface with a strong metal-to-metal interface (5). Other novel approaches such as Ru/Ti metal barrier (6), self-formed MnOx barrier (7), Ti metal barrier (8), and Ta implantation have also been reported (9).

Our investigation focused on the improvement of the SiCN/Cu interface, the metal barrier and metallization homogeneity. Integration results for PECVD Self Aligned Barrier (PSAB™), HCM IONX® Ta(N)/Cu process, HCM IONX Ti(N) metal barrier, novel Cu-Al alloy seed and PSAB/Cu-Al alloy seed synergies will be reviewed in this paper.

Experimental

A two-metal layer damascene structure with 65nm technology node features was used for the process integration study. HMS CORAL®, a dense low-k (k=3.0) SiOC film, was used as the inter-line dielectric. The barrier layer was deposited in a Novellus INOVA® system using HCM IONX Ti and HCM IONX Ta technology, respectively. Cu seed and

Ex. 1016, Page 297

Cu-Al seed were deposited in an INOVA system using HCM IONX Cu technology. The PSAB process was conducted on a VECTOR® Multi-Station Sequential Deposition (MSSD) platform. A via-fed electromigration test structure, with nominal line dimensions of 100 nm by 180 nm by 500 *u*m, was used for package EM testing. Via dimensions in the experiment are 90 nm, with nominal aspect ratio of 2:1, and current density range and measurement temperature are 1.5 – 3 MA/cm^2 and 275 – 325° C, respectively.

Results and Discussion

Metal Barrier and Seed Optimization

The metal microstructure and the interface along the liner and cap layer are both critical to interconnect reliability (10). Vias with high aspect ratios or "reentrant" profiles may lead to poor metal barrier coverage and the premature failure of the interconnect. Degradation of EM was observed when a non-uniform metal barrier process was used (3). A barrier-first approach with re-sputter was developed to provide good step coverage for the via bottom corner and the "reentrant" area of the feature (11).

Additionally, with thinner barriers film morphology and density become increasingly important to maintaining barrier integrity and ensuring a good interface with the copper seed. HCM IONX Ta(N) uses a high density plasma and optimal ion energy with Hollow Cathode Magnetron (HCM®) PVD technology to optimize the metal barrier microstructure. A high density metal barrier with smooth morphology and improved dielectric reliability was demonstrated in the experiment (12,13). Conformal copper seed step coverage is also essential to provide homogeneous metallization for the small features. RF copper deposition with resputter technology developed on the INOVA HCM source produces near-conformal copper step coverage. A combination of the HCM IONX Ta(N) and RF copper source leads to substantial improvements in Cu gapfill capability and EM performance. Significant improvements in both time-to-failure and sigma of distribution (Figure 1) are observed with the high density barrier and RF copper seed process (14).

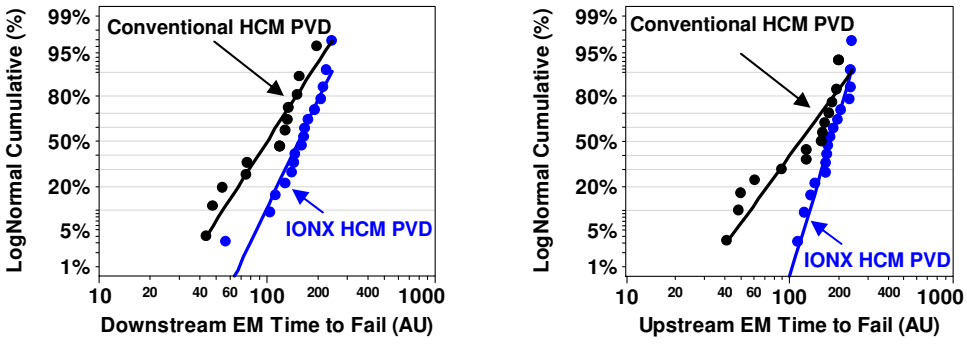


Figure 1. Effect of barrier/seed on EM: Significant improvement in both TTF and distribution is observed with the IONX PVD relative to conventional HCM PVD

PECVD Self-Aligned Barrier

The interface between SiCN and copper has been shown to be the main path for copper diffusion, and the weakest link in resisting EM failure. Many approaches have been explored to improve this SiCN/copper interface. Selective metal caps, such as

CoWP, have demonstrated significant EM improvement and received much interest. The implementation of a selective metal CoWP cap, however, is expensive and difficult due to limitations in achieving good selectivity to deposition over copper versus dielectric surfaces, and may lead to degraded dielectric reliability. A self-aligned barrier, on the other hand, has good deposition selectivity and may be implemented through a slight modification to the SiCN unit process without degrading the electrical isolation characteristics of the intra/inter-metal dielectric layers.

The PECVD self-aligned barrier (PSAB) has attracted much attention recently due to its high selectivity, low cost of implementation and benefits to interconnect reliability (15,16). The typical process steps in forming a PSAB have been described previously (17). High selectivity of the PSAB process arises through the natural differences in reactivity and reaction products of the gaseous constituents with the Cu and dielectric surfaces. In the case of Si-based PSAB, SiH_4 exhibits thermally-activated reaction with Cu, but the reaction on the dielectric surface results in the formation of an insulating film. Similarly, Ge reacts rapidly with Cu as shown in the SIMS analysis, but electrical testing shows no indication of Ge reaction with various dielectric films (17).

Significant improvement in mean-time-to-failure (MTTF) for the downstream EM test is noted with the integrated Ge-PSAB scheme as compared to the control (Figure 2). Details of the mechanism for EM improvement are still under investigation. Previously, we reported TDDB and EM improvements associated with improved adhesion of the low-k SiC to both Cu and low-k dielectric, enabling better encapsulation of the Cu inside the trench (15). In this work, adhesion energy measurements for both the control and Ge-PSAB/Cu interfaces were measured to be high and similar. The distribution of the Ge inside the Cu shows a higher concentration of Ge near the Cu-SiCN interface (17). The thin layers of $CuGe_x$ may aid in retarding the growth of the nascent defects (voids) within the interconnect structure. In addition, the cladding layers around the Cu surface create a shunt layer around the lower resistivity conducting metal, diverting the current away from the weaker interface. Diffusion of Cu with the electron wind would disturb the metallurgical equilibrium of the $CuGe_x$/Cu structure due to increases in the concentration of Ge in Cu-depleted areas near the interface. This creates a net force proportional to the concentration gradient, opposing electromigration(17). The above-mentioned mechanisms would lead to void growth delay without changing the failure mode. This is consistent with our investigation, where the EM activation energy of the Ge-PSAB is similar to that of reference SiCN, but the EM time-to-failure is much longer.

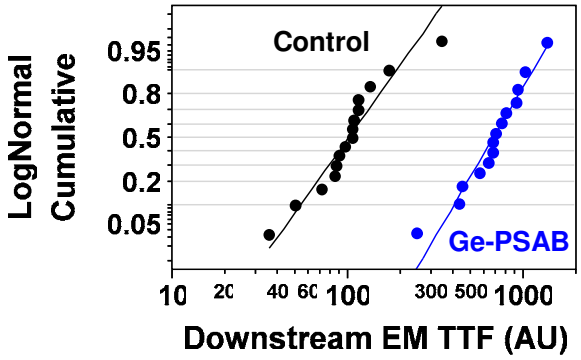


Figure 2. Downstream EM for the control and the Ge-PSAB splits

Novel Alloy Seed

Metal alloying has been widely used in semiconductor processes since the Al-interconnect era. Multiple Cu-alloying elements and approaches were investigated to improve reliability (18). Cu-Al alloys have received much attention. It was observed that alloying Cu with Al increases the incubation time for void growth and decreases the Cu drift velocity (19). PVD processing has been the preferred approach for Cu alloying due to the readily-available alloy targets, lower cost and easier process control. The alloying of Cu with Al, however, increases the Cu interconnect resistance and thus degrades RC delay. Our investigation of Cu-Al alloy seed integration focused on the tradeoff between increased resistance and improved reliability, with the goal of minimizing the RC delay degradation while maintaining the benefits of Cu-Al alloy seed. Through this process development and integration optimization, a novel alloy seed process was developed.

The downstream package electromigration test results suggest clear correlation between time-to-failure (TTF) and Al doping concentration for conventional alloy seed process (Figure 3). Significant EM improvement can be achieved with the Cu-Al alloy approach, but it leads to a large increase in RC delay. A measure of the tradeoff between EM improvement and an increase in RC delay, defined as the "EM Improvement Efficiency", was recently proposed by S. Yokogawa et al (20). Here it was concluded that with conventional Cu-Al alloy seed--while significantly improving EM-- the "EM Improvement Efficiency" was not as high as with the CoWP metal cap approach.

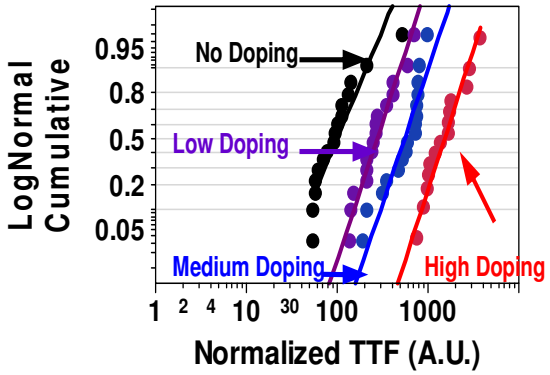


Figure 3. Downstream EM TTF as a function of Al doping concentrations for samples with conventional alloy seed.

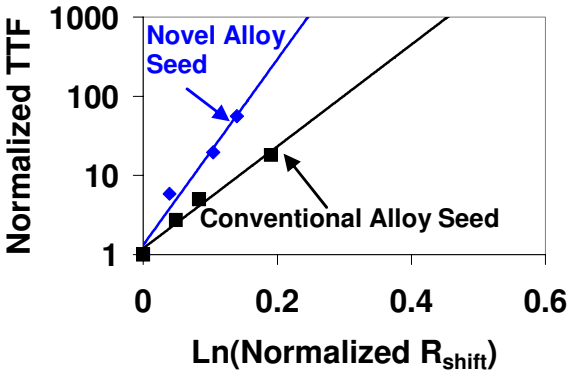


Figure 4. Comparison of the "EM Improvement Efficiency" between novel and conventional alloy seed process.

A novel alloy seed approach was developed with the HCM PVD process to enhance the "EM Improvement Efficiency" without compromising dielectric reliability and stress migration performance (12). The EM improvement with alloy seed as a function of the normalized line resistance is plotted in Figure 4. Consistently higher EM Improvement Efficiency, defined as the slope of the plot in Figure 4, is demonstrated with the novel alloy seed process. This novel process allows the use of a PVD process based on HCM technology to provide EM improvement for advanced technology nodes.

The improvement in EM with a Cu-Al alloy is attributed to a slower Cu drift velocity and a lower void growth rate (19). SIMS depth profile analysis indicates Al diffuses through Cu and accumulates at the SiCN/Cu interface upon thermal anneal (12). It is hypothesized that the Al getters the residual oxygen at the SiCN/Cu interface, driving the

segregation at the SiCN/Cu interface and improving electromigration. The improvement of the SiCN/Cu interface, along with slower Cu drift and reduced Cu void growth rate, greatly enhances electromigration performance with use of the Cu-Al seed. While the mechanisms are still under investigation, we believe that the novel alloy seed process allows the formation of a homogeneous Cu microstructure inside small features and strengthens the Cu/SiCN interface, leading to improved EM resistance.

Ti(N) Metal Barrier

The need for reliability improvement also led us to investigate alternative metallization approaches. A Ti(N) metal barrier was once considered to be inappropriate for Cu metallization due to potential F attack from fluorine-doped silicate (FSG). However, as carbon-doped oxide (CDO) replaces FSG for interconnect dielectrics, interest in Ti(N) barrier as the metal barrier has returned.

The interfaces between the Cu/Ti(N) metal barrier, Ti(N)/dielectric and SiCN/Cu are all critical to the reliability performance. Our dewetting investigation result suggests poor Cu wetting behavior on TiN, but excellent Cu wetting on Ti. However, Ti is known to diffuse into Cu, raising Cu resistivity, and is not able to provide the needed barrier properties. Here, a composite Ti/Ti(N)/Ti layer is proposed to address both Cu/Ti(N) interface and Cu diffusion barrier concerns (21).

No RC delay is observed with an optimal Ti(N) process, even though Ti diffusion into Cu is detected. This result can be attributed to accelerated Cu grain growth when Ti(N) is used as the metal barrier. Larger Cu grain alleviates the Cu resistivity increase due to Ti doping. SIMS profile analysis indicates Ti aggregation at the SiCN/Cu interface, similar to our observation with Cu-Al alloy seed. Ti also acts as an oxygen getter at the interface, and reinforces the SiCN/Cu interface. This mechanism is further supported by the higher adhesion energy and detection of TiOx at the SiCN/Cu interface. Significant EM improvement is observed with a Ti(N) metal barrier, as shown in Figure 5.

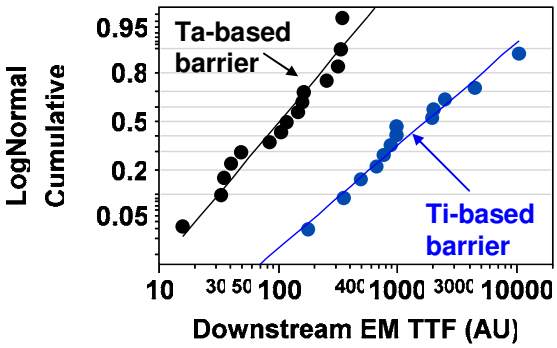


Figure 5. Downstream EM performance for Ta(N) and Ti(N) based metal barrier.

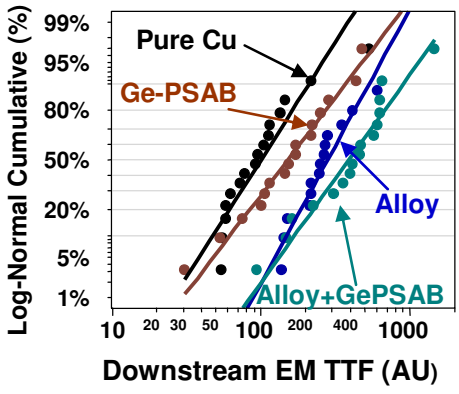


Figure 6. Progressive improvement in EM is demonstrated with Ge-PSAB, alloy seed, and Ge-PSAB + alloy seed.

Process Synergy

Multiple paths were explored for improvement of EM. A PSAB approach enhances the SiCN/Cu interface with the formation of a shunt layer near the SiCN/Cu interface,

Ex. 1016, Page 301

prolonging the voids growth within the interconnect structure. An alloy seed approach improves the EM performance through different mechanisms. The capability to combine various low cost, low risk reliability improvement approaches to meet the increasing reliability challenges provides greater benefits than employing radical process changes.

Figure 6 illustrates our preliminary result in exploring the synergistic effect of combining PSAB and alloy seed processes. The result shows improvement, whether with PSAB or Cu-Al alloy seed alone. Combining PSAB and Cu-Al seed leads to further improvement in EM time-to-failure. This synergy is not limited to PSAB and alloy seed, but is also applicable to other approaches. Recent studies suggest a similar synergy between the self-aligned barrier and the MnO barrier (22).

Conclusion

We have demonstrated multiple paths to improve EM performance. Optimization of the metal barrier and copper seed processes provides reliability improvement through a high quality metal barrier and homogeneous copper metallization. Further EM enhancement can be achieved through a novel copper alloy seed approach, which provides higher EM Improvement Efficiency over conventional Cu-Al alloy seed. The PSAB process shows significant EM improvement without compromising dielectric reliability. Furthermore, synergy between the Cu-Al alloy seed and PSAB demonstrates the capability to combine these different approaches to maximize the electromigration benefit. Finally, a significant EM improvement was observed with Ti(N) metal barrier integration, and RC delay was not compromised with the use of the Ti(N) metal barrier.

References

1. Y. Zhou, et al., *AMC,* (2004).
2. G.B. Alers, et al., *IITC,* (2003).
3. K.-D. Lee, et al, *IRPS,* 103 (2006).
4. A.H. Fisher, et al, *IRPS*, 511 (2007).
5. S. Gall, et al., *IITC,* 126 (2008).
6. M. Tagami, et al., *IITC,* 205 (2008).
7. T. Watanabe, et al., *IITC,* 208 (2008).
8. A. Sakata, et al., *IITC,* 101 (2006).
9. J. Gambino, et al., *IITC,* 22 (2007).
10. A.H. Fisher, et al., *IITC*, (2003).
11. G.B. Alers, et al., *IITC,* 2003.
12. H.J. Wu, et al., *Semiconductor International,* **11**, 38 (2008),
13. S.B. Law, et al., *AMC,* 133 (2007).
14. R. Shaviv, et al., *AMC,* 81 (2008).
15. K. Chattopadhyay, et al., *IRPS,* 128 (2006).
16. T. Usami, et al., *IITC,* 125 (2006).
17. H.-J. Wu, et al., *Semiconductor International,* **5**, 34 (2008).
18. K. Barmak, et al., *J. Vac. Sci. Technol. B,* **24(6)**, 2485 (2006).
19. S. Yokogawa et al., *J. Appl. Phys.,* **101**, 013513(2007).
20. S. Yokogawa, et al., *IEEE Transactions on Elec. Dev.*, **55(1)**, 350 (2008).
21 W. Wu, et al., *IITC*, 202 (2008)
22. H. Kudo, et al., *IITC,* 117 (2008).

Ex. 1016, Page 302

ECS Transactions, 18 (1) 275-280 (2009)
10.1149/1.3096462 © The Electrochemical Society

Use of MIR-FTIR and κ-value Measurements to Assess Potential Solutions to Reduce Damage During Porous Low-κ Integration

J. Beynet[a], D. De Roest[a], N. Rochat[b], K. Kellens[c], P. Verdonck[c], and H. Sprey[a]

[a] ASM Belgium N.V., Kapeldreef 75, B-3001 Leuven, Belgium
[b] CEA, LETI, MINATEC, F-38054 Grenoble cedex 9, France
[c] IMEC, Kapeldreef 75, B-3001 Leuven, Belgium

Multiple Internal Reflection (MIR) FTIR is used to assess the impact of potentially damaging BEOL integration process steps such as Chemical Mechanical Polishing (CMP) and NH_3-based plasmas on a κ=2.55 porous interlayer dielectric. The NH_3 plasma is found to be very damaging, and a He plasma prior to the same NH_3 plasma leads to a damage reduction. The impact of CMP is also confirmed and leads to an equivalent κ-value degradation as a He + NH_3 plasma.

In order to reduce the integration damage, the addition of a low-κ protective cap layer (κ=3.3) is proposed. The required thickness to avoid impact from CMP and NH_3-based plasmas is found to be 10 nm after CMP (κ_{stack}=2.60 with 10 nm cap). As a result, those process steps are no longer a concern with this integration route.

Introduction

Low-κ dielectrics have been introduced by the semiconductor industry in advanced integrated circuits interconnections in order to decrease signal delay, power consumption, and cross-talk. The main method to reduce the dielectric constant (κ) below κ=2.9 is to decrease the material density by the introduction of porosity. However, porosity tends to increase the susceptibility to damage of the low-κ materials during integration steps, such as Chemical Mechanical Polishing (CMP) (1) and NH_3-based plasmas performed before the dielectric barrier deposition (2). Those plasmas, applied here on blanket low-κ layers, aim at reducing the copper oxide, nitridating the subsequently cleared copper surface and also removing CMP slurry residues from the low-κ. During a standard integration flow the CMP consists of three steps: the main copper removal step, the remaining selective copper removal on top of the Ta/TaN barrier, and the Ta/TaN barrier removal step. Since the low-κ layer is not exposed to the two first CMP steps, the CMP process studied here is the Ta/TaN barrier removal step, using a slurry non-selective to the low-κ. This step is performed on blanket low-κ material, without Ta/TaN barrier on top. Blanket layer studies are appropriate to evaluate the damage caused by those process steps, since they are typically impacting the low-κ material from the top surface in an integrated structure.

Experimental

The dielectric material for this study is deposited by Plasma Enhanced Chemical Vapor Deposition (PECVD) in an ASM® Eagle® 12 300 mm chamber, resulting in a thin

Ex. 1016, Page 303

film composed of an organic porogen phase embedded in an a-SiOC:H matrix. A subsequent UV curing process removes the porogen, after which a porous material is created (κ=2.55). The protective cap layer is deposited in the same chamber. The NH_3-based plasma is conducted in an ASM® Eagle® 12 chamber and normally performed *in-situ* with the subsequent dielectric barrier deposition (not studied in this paper) in a regular integration flow.

The FTIR spectroscope is a Nicolet 6400 for the transmission mode analysis (4 cm^{-1} resolution, 128 scans accumulation, 4000-400 cm^{-1} range) and a Bruker IFS 55 for the Multiple Internal Reflection (MIR) mode (4 cm^{-1}, 200 scans, 5000-1500 cm^{-1}). The MIR set-up permits to drastically increase (300 times) the optical path length inside the film and thus the absorption intensity (3) (4). Thickness and refractive index were measured at 633 nm by spectroscopic ellipsometry with a KLA Tencor F5 tool. C(V) curves were measured using a Metal Insulator Semiconductor (MIS) structure with an MDC Mercury probe and an HP 4156C LCR meter at 0.1 MHz. Highly doped wafers are used in order to constantly be in the accumulation regime. The capacitance obtained was used to calculate the dielectric constant. The material's hydrophobicity was determined by contact angle measurement of de-ionized water (18.2 MΩ.cm) on the film surface by using a Data Physics SCA20 contact angle analyzer.

Results and Discussion

Damage characterization

NH_3-based plasmas.

The thin low-κ layer has been exposed to three different plasma processes. The first one is the conventional process using NH_3, the second is a pure He plasma and the third is the sequence of the He and the NH_3 plasma. For the NH_3 plasma treatment only, the thickness and the refractive index slightly decrease, respectively by 3.6% (-7.5nm) and 0.44% (-0.006) (Table I). Furthermore, the top surface went from hydrophobic to hydrophilic, with a water contact angle drop from 93º to less than 5º after plasma. The most important impact is the κ-value increase by +44%, from 2.57 to 3.69. It is clearly illustrated that the ammonia plasma is very damaging for the porous low-κ material.

A standard transmission FTIR analysis of the thin layers (Figure 1) shows no drastic impact on the Si-O-Si and Si-CH_3 peaks. The FTIR Multiple Internal Reflection (MIR) mode is used in order to enhance the absorption intensity in the 1500-4000 cm^{-1} range. Figure 2 highlights the pronounced signal increase in this area. The comparison of the MIR spectra of the materials before and after NH_3 plasma shows an absorption increase in the 3000-3800 cm^{-1} range, indicating an –OH group creation. This chemical group generation is consistent with the κ-value increase. Additionally, there is a slight decrease of Si-H groups, but the amount of CH bonds (in CH_2 and CH_3) remains identical. It perfectly illustrates the benefits of an MIR analysis, in addition to the standard transmission mode, for investigating accurately the –OH, -CH and Si-H bonds.

A proposed process to decrease the damage of the NH_3 plasma on the porous low-κ is to first perform a He plasma, prior to the same NH_3 plasma. The proposed mechanism is a densification of the top layer, causing a pore sealing which reduces the plasma diffusion inside the material.

First, the material impact of the He plasma only is assessed. The thickness slightly decreases by 2.8% (-5.8nm) and the refractive index increases by 0.37% (+0.005),

Ex. 1016, Page 304

consistent with a densification associated to a shrinkage (Table I). The transmission FTIR analysis in figure 1 shows no observable impact on the Si-O-Si and Si-CH_3 peaks, but MIR illustrates a moderate –OH increase, consistent with the slight rise in κ-value (+2%, from 2.57 to 2.62). So, the overall low-κ film properties seem to be only moderately impacted by the He plasma process, but the contact angle drop from 93º to less than 53º suggests a chemical modification concentrated at the top surface.

Then, a combination of the two plasma processes is evaluated on the pristine low-κ material: a sequence of the He plasma and the NH_3 plasma. In comparison with the NH_3 plasma only, a similar thickness and RI decrease (respectively -3.6% and -0.009) and hydrophilicity (contact angle < 5º) is observed. However, the κ-value increase is drastically reduced when using a He plasma prior to the NH_3 plasma: +21% instead of +44%. The MIR spectra comparison highlights an analogous chemical impact with and without a prior He plasma, but to a lesser extend in the first case: –OH group creation, slight decrease of Si-H groups, and identical CH bonds amount. The lower –OH generation correlates well with the lower κ-value increase.

TABLE I. Plasma impact on low-κ material.

Material	After low-κ deposition		After plasma			
	Thickness (nm)	RI	Thickness (nm)	RI	k-value	Water contact angle
low-κ (reference)	206.0	1.363	NA	NA	2.57 (NA)	93º
Low-κ + NH_3 plasma	206.0	1.363	198.5	1.357	3.69 (+44%)	<5º
Low-κ + He plasma	205.6	1.363	199.8	1.368	2.62 (+2%)	53º
Low-κ + He & NH_3 plasma	205.7	1.363	198.3	1.354	3.10 (+21%)	<5º

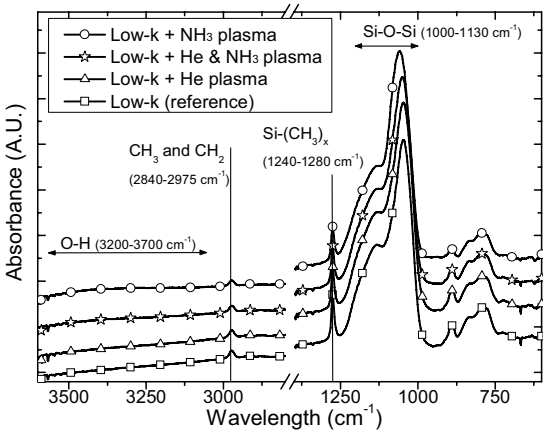


Figure 1. Transmission FTIR spectra of the low-κ material before and after plasma

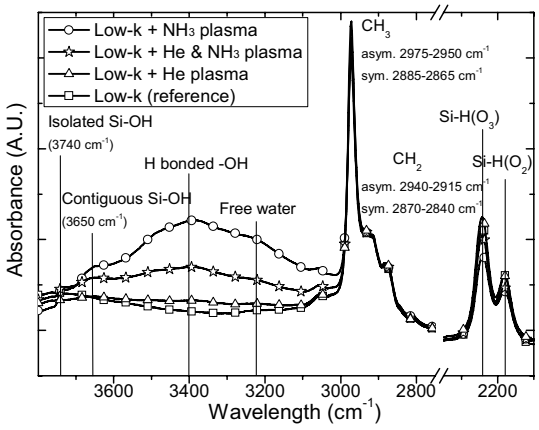


Figure 2. MIR-FTIR spectra of the low-κ material before and after plasma

Chemical Mechanical Polishing.

Following a similar procedure, the CMP impact on the κ=2.55 material was assessed (Table II). After the CMP process, the thickness decreases by 10.6 nm, the refractive index increases from 1.363 to 1.404, and the κ-value drastically increases by 21%, consistent with a removal by polishing coupled with a densification. The MIR spectra subtraction (after – before CMP) in figure 3 highlights the difference of absorption due to the CMP process: a clear CH_2 group generation and a slight –OH group increase.

TABLE II. CMP impact on low-κ material.

Material	After low-κ deposition		After CMP		
	Thickness (nm)	RI	Thickness (nm)	RI	k-value
low-κ (reference)	206.0	1.363	NA	NA	2.57
low-κ + CMP	204.0	1.363	193.4	1.404	+21%

The described experiments illustrate the impact of NH_3-based plasmas and CMP on a 2.55 porous low-κ material. The NH_3 plasma is found extremely damaging and a prior He plasma process is demonstrated to reduce, but not prevent, this impact. Even a short CMP step is found as damaging as this He + NH_3 plasma process. As a result, all the benefit of the porosity creation to reduce the κ-value is lost during integration.

Integration route to limit damage

In order to circumvent this issue, the addition of a protective cap layer (κ=3.3) on top of the low-κ is proposed. A dense layer could limit or prevent the diffusion into the porous structure of liquid chemicals during the CMP process and of reactive species during the plasma process.

NH_3-based plasmas.

The cap layer addition is first assessed with respect to the NH_3-based plasma process. The initial target thickness is 10 nm, a good trade-off between moderate increase of the stack overall κ-value and controllability of the deposition process. Despite the more challenging ellipsometric measurement on a bi-layer structure after cap layer deposition, measurements on the four wafers show the same cap layer value: 8.3 nm (Table III). The slight κ-value and RI increase of the stack after cap deposition is consistent with the higher values of the dense material (respectively 3.3 and 1.45). The κ-value increase of the stack after the most damaging process studied in this paper, the NH_3 plasma, is marginal (only 0.03) and is even equal to zero with a prior He treatment. This result is confirmed by MIR-FTIR analysis: no significant –OH group creation is observed in the stack after the plasma processes, as illustrated by the spectra subtraction to the initial layer (Figure 4). The He plasma effect on the stack is comparable to the impact previously observed on the low-κ only (Table I), and is consistent with a densification.

A thinner cap layer could even be tested, but 10 nm is thought to be preferable for the CMP process robustness. It is obvious that a thicker cap layer has to be deposited, because part of it is removed during the metallic barrier clearance with a non-selective slurry by CMP.

TABLE III. Plasma impact on stack (low-κ + 10 nm cap).

Material	After low-κ deposition		After cap deposition		After plasma		
	Thickness (nm)	RI	Thickness (nm)	RI	Thickness (nm)	RI	k-value
Low-κ (reference)	206.0	1.363	NA	NA	NA	NA	2.57
Low-κ+cap (10 nm)	206.6	1.363	214.9 (+8.3nm)	1.364	NA	NA	2.60
Low-κ+cap+NH_3 plasma	206.9	1.363	215.2 (+8.3nm)	1.364	213.4 (-1.8nm)	1.365 (+0.001)	2.63
Low-κ+cap+He&NH_3 plasma	207.2	1.362	215.5 (+8.3nm)	1.364	210.5 (-5.0nm)	1.371 (+0.007)	2.60
Low-κ+cap+He plasma	206.8	1.363	215.1 (+8.3nm)	1.364	210.4 (-4.7nm)	1.373 (+0.009)	2.60

Ex. 1016, Page 306

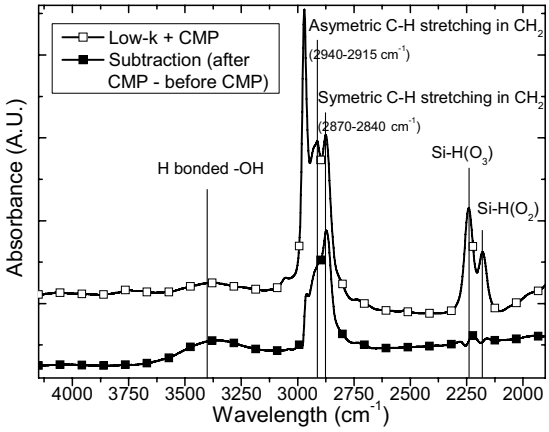


Figure 3. MIR-FTIR spectra of the low-κ after CMP and subtraction to the reference

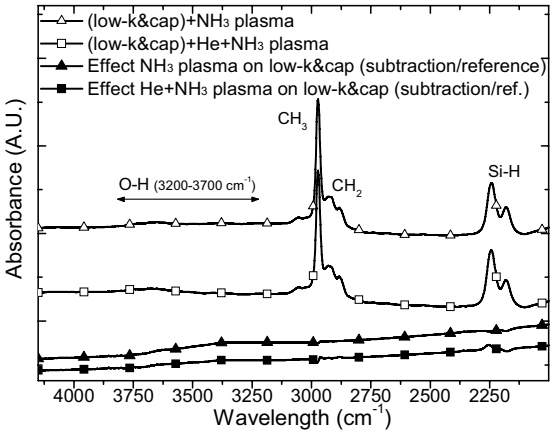


Figure 4. MIR FTIR spectra of the (low-κ + 10nm cap) after NH_3 and He + NH_3 plasmas

Chemical Mechanical Polishing.

In order to estimate the thickness of cap material removed during CMP and to investigate the intrinsic material modification during this process, a 140 nm cap layer is polished (Table IV). The RI is almost not changed by CMP (+0.002) and 22.8 nm are removed. Furthermore, neither the κ-value nor the MIR-FTIR chemical signature are significantly impacted (Figure 5). The cap material is found to be an appropriate layer to withstand the CMP process.

The same CMP process is performed on the following stacks: low-κ plus a 10, 25 and 50 nm cap layer (Table IV). The CMP impact with respect to the κ-value is still observed with a 10 nm cap layer (+10.3%) but reduced by a factor two compared to the impact without cap. The RI increase is also reduced: +0.018 instead of +0.041. The κ-value is no longer significantly impacted with 25 nm cap material (only +0.4%). MIR-FTIR spectra display a similar trend (Figure 6). Indeed, the CH_2 uptake is reduced with the 10 nm cap layer, and even not significant anymore with the 25 and 50 nm caps. A comparable observation is made with the –OH group in the 3200-3500 cm^{-1} range. So an additional cap layer is effective in protecting the porous low-κ material against CMP-related chemical modifications and deterioration of the κ-value. The optimum cap thickness is between 10 and 25 nm. Because approximately 20 nm cap is removed by the CMP process on blanket wafers, it is believed that this thickness (or slightly above for robustness of the integration scheme) is needed to avoid CMP impact on blanket films.

TABLE IV. CMP impact on low-κ + cap layer.

Material	After low-κ deposition		After cap deposition		After CMP		
	Thickness (nm)	RI	Thickness (nm)	RI	Thickness (nm)	RI	k-value
Cap (reference)	NA	NA	153.1	1.447	NA	NA	3.30
Cap + CMP	NA	NA	153.2	1.447	130.4	1.449	+0.8%
low-κ (reference)	206.0	1.363	NA	NA	NA	NA	2.57
low-κ + CMP (reminder)	204.0	1.363	NA	NA	193.4	1.404	+21%
Low-κ + 10 nm cap + CMP	205.7	1.363	214.1	1.364	197.7	1.381	+10%
Low-κ + 25 nm cap + CMP	205.1	1.363	230.4	1.375	214.2	1.364	+0.4%
Low-κ + 50 nm cap + CMP	203.6	1.364	256.2	1.434	247.4*	1.407*	-1.1%*

* Goodness of fit very low for this measurement, attributed to a less accurate thickness measurement because of the combination of a stacked sample and a top layer gradient. This thickness measurement problem directly impacts the k-value measurement.

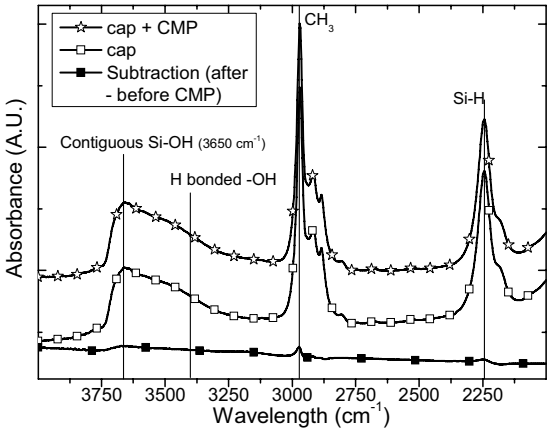


Figure 5. MIR FTIR spectra of the cap material before and after CMP

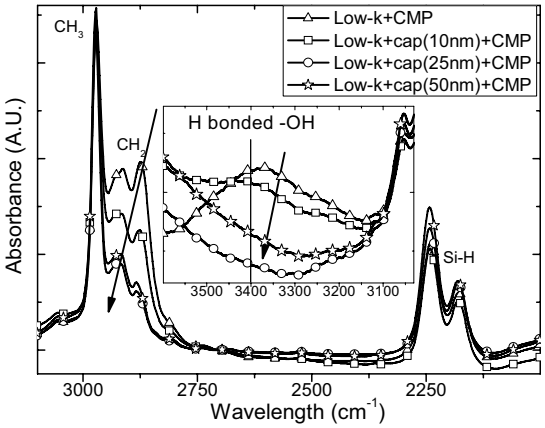


Figure 6. MIR FTIR spectra after CMP of different stack (low-κ + cap layer)

Conclusions

Impact of Chemical Mechanical Polishing (CMP) and NH_3-based plasmas has been assessed on a κ=2.55 porous dielectric material. Both processes damage the porous material and can respectively increase the κ-value up to 3.1 and 3.7. MIR-FTIR is found very useful to characterize the chemical impact: mainly CH_2 bonds creation during CMP and –OH groups generation by NH_3-based plasmas.

The proposed solution to eliminate this negative impact is the addition of a protective dense low-κ cap layer (κ=3.3) on top of the porous film. This study demonstrates the cap material ability to withstand the CMP process and establishes the required thickness to avoid CH_2 generation and κ-value increase: 25 nm. Furthermore, a 10 nm cap layer prevents properly the -OH group creation and associated κ-value rise upon exposure to a He + NH_3 plasma. Consequently, a 35 nm cap layer deposition on top of the κ=2.55 film could result in an overall κ=2.60 stack, after CMP and NH_3-based plasma. An efficient route is proposed to solve integration damage.

Acknowledgments

The authors would like to acknowledge N. Kobayashi and K. Matsushita from ASM Japan K.K. for support and discussion.

References

1. A. Castex, S. Gall, J. Farkas, P. Holliger and M. Rivoire, in *12th international C.M.P. planarization for ULSI multilevel interconnection conference*, session III CMP reliability issue (2007).
2. N. Posseme, T. Chevolleau, T. David, M. Darnon, O. Louveau and O. Joubert, *J. Vac Sci Technol. B*, **25**, 1928 (2007).
3. N.J. Harrick, Internal Reflection Spectroscopy, Wiley, New York, 1967.
4. N. Rochat, A. Troussier, A. Hoang, and F. Vinet, Mat. Sci. Eng., **C 23**, 99 (2003).

Ex. 1016, Page 308

ECS Transactions, 18 (1) 281-286 (2009)
10.1149/1.3096463 © The Electrochemical Society

Performance Improvement of CuOx RRAM for Non-volatile Applications

Y.Y. Lin[a*],P. Zhou[a*], M. Yin[a], H. B. Lv[a], S. Song[b], R.Huang[b], J.G.Wu[b] , H. M. Wu[b] and M. H. Chi[b]
[a] State Key Lab of ASIC & System, Research Center of Semiconductor Memory and Application, Fudan University, Shanghai 200433, China.
[b] Memory Technology Development Center, Semiconductor Manufacturing International Corp., Shanghai, 201203, China
*Corresponding author: yylin@fudan.edu.cn, pengzhou@fudan.edu.cn

We found CuOx thin film with gradual oxygen concentration (GOC) distribution enhances resistive switching characteristics for nonvolatile memory applications. Using Al/ GOC CuOx /Cu structure, not only no forming is needed, also the endurance of switching is greatly enhanced. The device with GOC CuOx demonstrates resistance on/off ratio greater than 100, and endurance of more than 12000 cycles. A ramped pulses series (RPS) operation method is also put forward, which can improve switching stability and cycling endurance remarkably. A method for minimizing the dispersion of Vreset by optimizing amplitude of pulse is proposed further. Dispersion of Vreset can be minimized by using optimized Vstart of RPS. And this phenomenon agrees well with the Joule heating model that the RESET process of memory switching is associated with rupturing of filaments by Joule heating.

Introduction

The Binary oxide-based resistive random access memory (RRAM)1 has attracted growing interest as one of the most promising alternatives for easing challenges of flash scaling down, owing to its good scalability and ease of fabrication[1-3]. But several issues have to be solved for practical application, including instability of operation, endurance, data retention, integration etc. CuxO based RRAM has high affinity with standard logic process. The good resistive switching characteristics of Cu_xO has been reported, where the maximum reported endurance of Cu_xO-based RRAM cell is about 1000 cycles [4-5].

In this work, we report that the copper oxide based RRAM cell can achieve cycling up to 12000 cycles as well as no initial "forming process" needed. This "forming-free" characteristic of device appears related to thinner CuO cap layer with saturated oxygen; and the endurance enhancement appears related to the profile and magnitude of oxygen vacancies in CuxO[6-7]. Meanwhile, stable switching characteristics and narrow distribution of resistive switching parameters are improved.

Experiment

The device (Al/ GOC CuOx/Cu) can be prepared by electro-chemical plating (ECP) surface treatment by plasma reactive ionic etch (RIE), and oxidation by inductively coupled plasma (ICP) or thermal process. Cu film with the thickness of 1 μm was deposited on Ta/TaN/SiO2/Si substrate with standard Novellus® ECP system at room temperature. After Cu film was deposited, the surface is slightly etched (for removing native Cu oxide on surface) by RIE and then plasma oxidized in ICP chamber. The rest of Cu layer acts as the bottom electrode. 180 × 180μm Al top electrodes with the thickness of 400 nm were fabricated by evaporation and lift off process. Keithley 4200-SCS

semiconductor parameter analyzer was used to measure the current–voltage (I–V) characteristics. Keithley 3402 pulse pattern generator was used to generate voltage pulses. All the measurements were performed on a manual probe station at room temperature.

Results and discussion

Figure 1 shows a typical structure of the Al/ GOC CuOx/Cu device in this study and its profile of oxygen concentration in CuOx as derived from the depth-Auger electron spectroscopy (AES) measurements. The inset of Fig. 1 sketches the Cu-oxide (total ~150nm thick) possibly separated into two layers with the upper mainly as approximate CuO layer without clear interface (~60nm) and the lower as CuOx with gradual decreasing of oxygen concentration. The AES signal in Fig. 1 represents a relative population of oxygen atoms with respect to total Cu and O atoms, where the initial 50% of O-atoms simply indicate a composition of CuO (i.e. x~1) near the Al electrode for ~60nm (as estimated from the sputtering time). The oxygen concentration starts decreasing until reaching 0% to pure Cu at bottom, where the slope -5.88 represents the gradient of oxygen profile. As the oxygen concentration decreases toward the bottom Cu electrode, there are increasingly oxygen vacancies rich. Electrically, those oxygen vacancies are positively charged and may behave as holes, as consistent to the known Cu_2O as p-type semiconductor with holes as charge carriers. Thus, the lower layer of CuOx in device is oxygen vacancies rich and more conductive (i.e. lower resistivity). Similarly, the upper layer of CuO near the Al electrode is less conductive (i.e. higher resistivity) due to much less oxygen vacancies.

The thickness of the upper approximate CuO layer has significant effect on the initial “forming” process as shown in the inset of Fig.2. In contrast, the Al/GOC CuxO/Cu devices with thinner (~60nm) and thicker (~170nm) upper CuO layer need ~2V and ~9V bias respectively for initial triggering the Al/ GOC CuxO/Cu device from the high resistance state (HRS, off-state) to low resistance state (LRS, on-state) (i.e. SET operation) as in the inset of Fig. 2. After the initial “forming” by the 1st SET operation, the trigger voltage is lowered to ~2v in subsequent SET operations as in Fig.2. Thus the thinner upper approximate CuO layers in Al/CuxO/Cu device practically results in “forming-free”, while a thicker upper approximate CuO resulting in “forming-necessary” devices. On the other hand, the upper approximate CuO layer shall be thick enough to avoid Fowler-Nordheim tunneling leakage current during SET operation. Thus, the thickness of the upper CuO layer is a design parameter for eliminating the initial “forming” process, enhancing the resistance on/off ratio, and improving the uniformity of on-state resistance. It is an effective device parameter for optimizing performance of CuxO based RRAM.

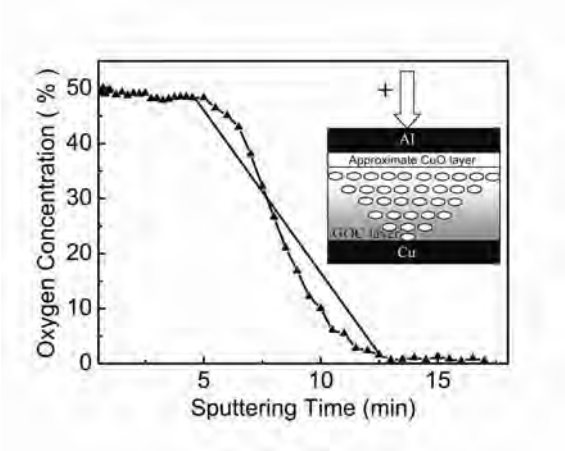


Figure 1. The depth-distribution profile of oxygen concentration in CuOx film for this RRAM device. The dash line is fitted line of gradual oxygen concentration region after approximate CuO layer, and the slope is -5.88. The inset figure indicates schematic oxygen concentration distribution and structure of RRAM device.

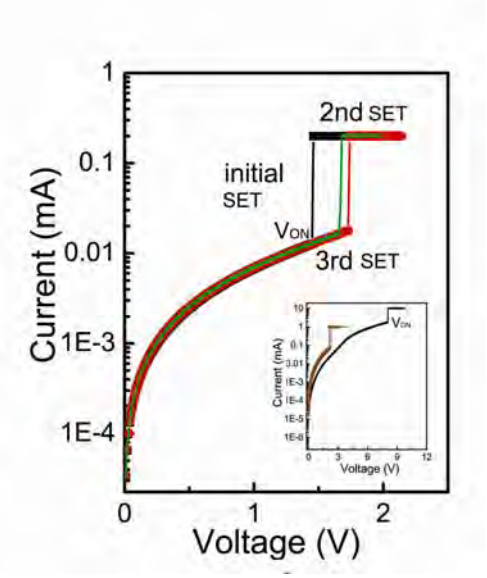


Figure 2. I-V characteristics of the RRAM device cell in SET process with 180 diameter Al top electrode.

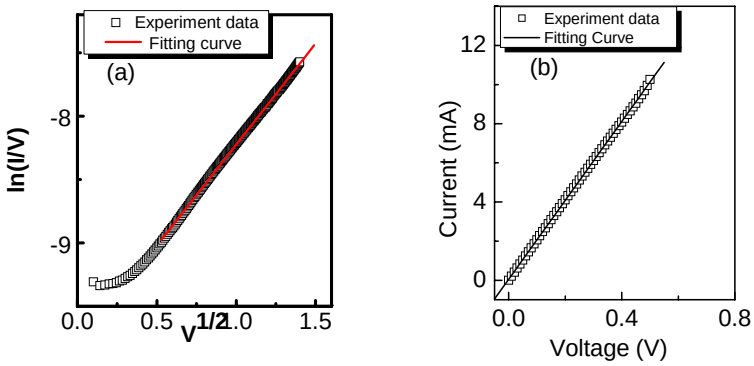


Figure 3. (a). The logarithmic plot of I versus of HRS. (b). The I-V curve of LRS. The solid lines signify the fitting curve of experiment data.

Frenkel-Poole emission model is extensively used to investigate carrier transport behavior in insulator and semiconductor, which expression can be described by

$$I \propto V \exp(2\alpha\sqrt{V}/T - q\varphi b/KT)$$

Where I is the current, q is the electron charge, V is the applied voltage, α is a constant, φb is the energy barrier height, k is the Boltzmann constant, and T is the Kelvin temperature, respectively. Fig. 3 (a) shows the logarithmic plot of I versus $\sqrt{V}$ of HRS. The experiments data obeys a good linear relationship at high electrical field region. From the $\ln I/V \propto \sqrt{V}$ relation and Frenkel-Poole constant α calculation, carrier

trapping and de-trapping of Frenkel-Poole effect is thought as the main conduction mechanism of HRS. Fig. 3 (b) shows the *I-V* curve of LRS or on state. It can be noticed that the current has a linear relationship with voltage of on-state, which indicates that Ohmic conduction is dominant in LRS. And it suggests conductive filament paths are formed as a bridge between top electrode and bottom electrode after SET process.

In order to avoid the RESET instability and minimize the Vreset dispersion in pulse operation mode, a ramped pulses series mode(RPS) are proposed for RESET operation for the first time as shown in Fig. 4. It's a kind of write-verify programming algorithms. The RPS includes a series of pulses. These pulses ramp from initial voltage (Vstart) to last voltage (Vend) with a fixed voltage step (Vstep). All the single pulses of RPS are same in time width but different in voltage amplitude. The Vend is determined by maximum Vreset. A resistance read process is performed after each single pulse. Once the resistance reaches reference value of HRS, the RESET process will be terminated and cancel the rest pulses of RPS. Fig. 5(a)-4(d) are statistical results of Vreset in 1000 switching cycles when using RPS mode for RESET. Four charts show the dispersion of Vreset vary with Vstart, while Vstep, Vend and SET pulse are fixed. When RPS started from -0.5V, the probability of achieving RESET at first pulse (0.5V) is 66.6% (Fig. 5(a)), but it needs RPS extending to -1.3 V (the fifth pulse) to ensure complete RESET success. When RPS started from -0.7 V, the device can be successfully switched to HRS at first pulse in 98.4% cycles, and the rest can be achieved at second pulse (Fig. 5(b)). However, when the Vstart is continually increased, the probability of achieving HRS at first pulse begins to decrease and Vreset dispersion becomes higher with a maximum Vreset of -1.3 V and -1.7 V needed respectively as shown in Fig. 5(c) and (d). It confirms that the maximum Vreset can not be adopted as the amplitude of RESET pulses and there is an optimized Vstart that can minimize the dispersion of Vreset for the device. Fig.5(e) show that the peak power decreases as temperature increases which indicates the thermal benefit of RESET transition and supports Joule heating RESET model.

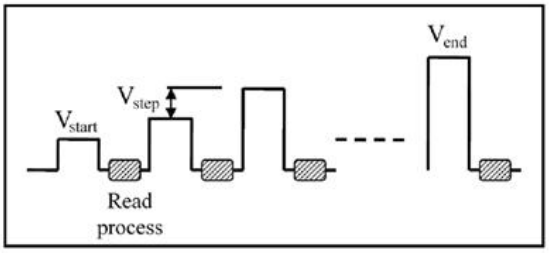


Figure 4. Schematic sequence of RPS for RESET operation.

Ex. 1016, Page 312

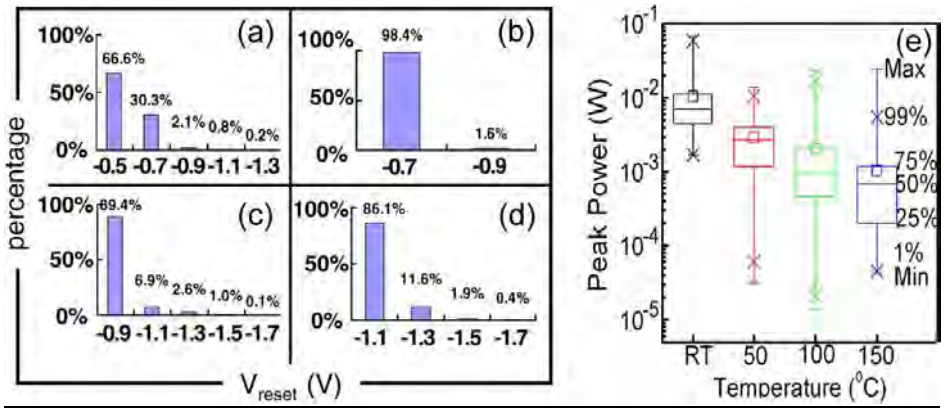


Fig.5 (a) - (d) Percentage of Vreset in RPS RESET mode with different Vstart: (a) -0.5 V, (b) -0.7 V, (c) -0.9 V, (d) -1.1 V. The Vstep and Vend is -0.2 V and -2 V respectively, pulse width is 10 ms. SET by pulse of 5 V for 1 ms. (e) Statistical charts of peak power of RESET transition during DC voltage sweeping under various temperatures. RT denotes room temperature. Unfilled square indicates the mean value of the distribution.

Figure 6 shows the endurance performance of Al/ GOC CuOx/Cu device with thin (~60nm) upper approximate CuO layer by means of RPS reset process. The endurance of 12000 times switching cycles is achieved.

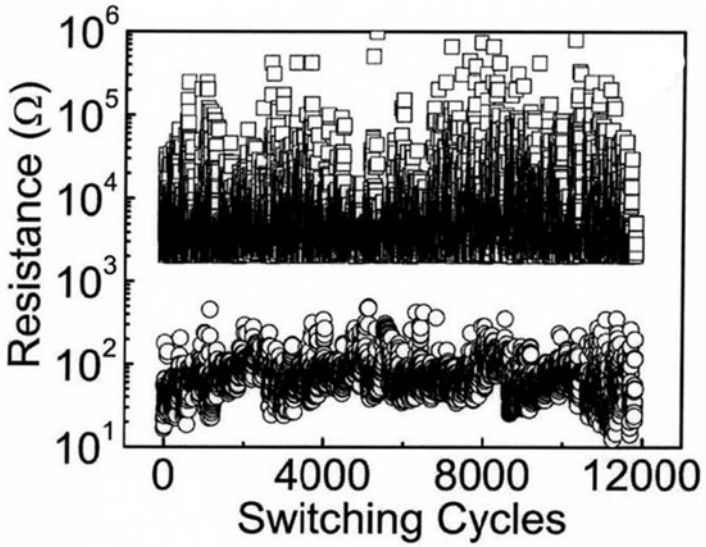


Figure 6. Superior 12000times switching cycles and clear operation windows of Al/GOC CuOx /Cu device .

Summary

In summary, CuOx with GOC shows very promising improvement in endurance and forming-free characteristics under the RPS mode reset operation. The fabrication of the Al/ GOC CuOx /Cu structure is fully compatible with standard CMOS as a strong advantages compared to other technologies. The switching behavior is consistent with both the Poole-Frenkel effects in HRS and the formation/rupture of conductive multi-filaments in CuOx during switching. This phenomenon of the first-pulse altitude and temperature dependence agrees well with the Joule heating model that the RESET process of memory switching is associated with rupturing of filaments by Joule heating.

Ex. 1016, Page 313

Acknowledgments

This work was supported by the NSFC (60706033 and 60676007), 863 project of China (2008AA031401) and 973 project of China (2007CB935403).

References

1. An Chen, Sameer Haddad, Yi-Ching (Jean) Wu, Tzu-Ning Fang, Zhida Lan, Steven Avanzino, Suzette Pangrle, Matthew Buynoski, Manuj Rathor, Wei (Daisy) Cai, Nick Tripsas, Colin Bill, Michael VanBuskirk, and Masao Taguchi, IEDM Tech. Dig., 2005, pp.746-749.
2. B. J. Choi, D. S. Jeong, S. K. Kim, C. Rohde, S. Choi, J. H. Oh, H. J. Kim, C. S. Hwang, K. Szot, R. Waser, B. Reichenberg, and S. Tiedke, J. Appl. Phys., vol. 98, no. 3, p. 033 715, Aug. 2005.
3. X. Wu, P. Zhou, J. Li, L. Y. Chen, H. B. Lv, and Y. Y. Lin, Appl. Phys. Lett., vol. 90, no. 18, p. 183 507, May. 2007.
4. T.N. Fang, S. Kaza, S. Haddad, A. Chen, Y.C. Wu, Z. Lan, S. Avanzino, D. Liao, C. Gopalan, S. Choi, S. Mahdavi, M. Buynoski, Y. Lin, C. Marrian, C. Bill, M. VanBuskirk and M. Taguchi, International Electron Devices Meeting Tech. Digest, San Francisco, 439 (2006).
5. R. Dong, D. S. Lee, W. F. Xiang, S. J. Oh, D. J. Seong, S. H. Heo, H. J. Choi,M. J. Kwon, S. N. Seo, M. B. Pyun, M. Hasan, and H.S Hwang, Appl. Phys. Lett. 90,042107 (2007).
6. M .Yin, P. Zhou , H.B. Lv , J. Xu, Y.L Song , X.F.Fu , T.A.Tang , B.AChen and Y.Y.Lin , IEEE Electron Device Letters. 29, 681-683, (2008)
7. P. Zhou, H.B. Lv, M. Yin., L. Tang , Y.L.Song, T.A.Tang, Y.Y.Lin , A. Bao, A. Wu, S. Cai, H. Wu, C. Liang, and M. H. Chi Journal of Vacuum Science & Technology B, 26,1030-1033, （2008）

ECS Transactions, 18 (1) 287-290 (2009)
10.1149/1.3096464 © The Electrochemical Society

Yield Improvement through Mechanism Study on WSi_2 Abnormal Growth

Ji Yue Tang, Zhijun Fang, Jerry Xu

Applied Materials China

Abstract

More than 7% yield loss from a 0.11 um deep trench DRAM product line was incurred due to short current between the gate contact and the bit line contact (GC CB short) caused abnormal growth of WSix. Tungsten rich during WSi_2 deposition was suspected by customer at the beginning of the issue**.** Further study shows that the oxygen attacking during GC annealing stage is the culprit which causes the abnormal growth of WSix. The paper also explains in detail on why WSi_2 will grow abnormally after the oxygen attack. The paper also explains how can the oxygen attack the WSi_2 layer when there is a silicon nitride cap layer on top of the WSi_2 film. There is still one more question remain, why O_2 in GC RTO process which is just right after GC anneal process will not cause WSi_2 abnormal grow? As a summary, the O_2 control in GC anneal step is the key to eliminate the WSi_2 tumor defects. GC anneal recipe was modified to avoid O_2 sensor delay effect to keep the O2 under control.

Introduction

7% yield loss due to GC CB short caused by WSix tumor. Tungsten rich during WSi_2 deposition was suspected by customer**.** Further study shows that the oxygen during GC annealing is the culprit to cause the WSi_2 abnormal grow and cause the yield loss. The paper explain the details on why WSi_2 will grow abnormally after oxygen attack, how can the oxygen attack WSi_2 even there is silicon nitride cap layer on top of the WSi_2 film, why O_2 in GC RTO process which just right after GC anneal process will not cause WSi_2 abnormal grow.

Approach and Study

The atomic lattice constant of W_5Si_3 is much bigger than $WiSi_2$ (see table 1). If WSi_2 phase is transformed to W_5Si_3, the WSix will grow abnormally.

Ex. 1016, Page 315

Table 1. Lattice constant of different WSix phases

	Crystal structure	Lattice constant (au)
WSi_2	Tetragonal C11b	a = 6.0698 c = 14.891
	Hexagonal C40	a = 8.7192 c = 12.121
W_3Si_2	Tetragonal D8m	a = 18.072 c = 9.3384
W_5Si_3	Hexagonal D8	a = 13.587 c = 9.1952
	Tetragonal D8m	a = 18.151 c = 9.3806
W_3Si	Cubic cP8	a = 9.2786

Figure 1 shows the two element phase diagram, the phase of W_5Si_3 can be formed only if W is very rich during WSi_2 deposition. A lot of works were done to check if there is W rich during WSi2 deposition. There is no obvious evidence of W rich was found.

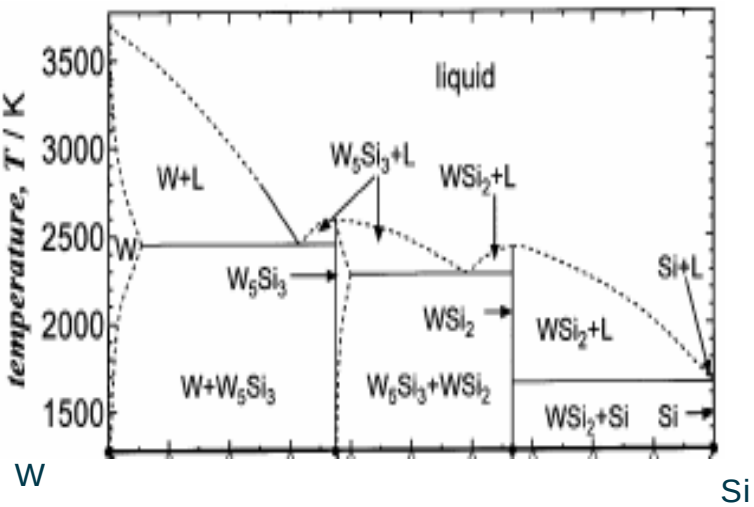


Figure 1 Two element phase diagram of W and Si.

Nitride deposition, gate etch and GC anneal processes are studied in detail for finding the root cause. Fig 2 shows the 3 element phase change diagram, oxygen will react with WSi_2 and form W_5Si_3 through below equation even there is no W rich,

$$\mathbf{5\ WSi_2 + 7\ O_2 = W_5Si_3 + 7\ SiO_2}$$

Thus W_5Si_3 phase will be formed when WSi_2 was attacked by oxygen even there is no W rich in WSi_2 phase.

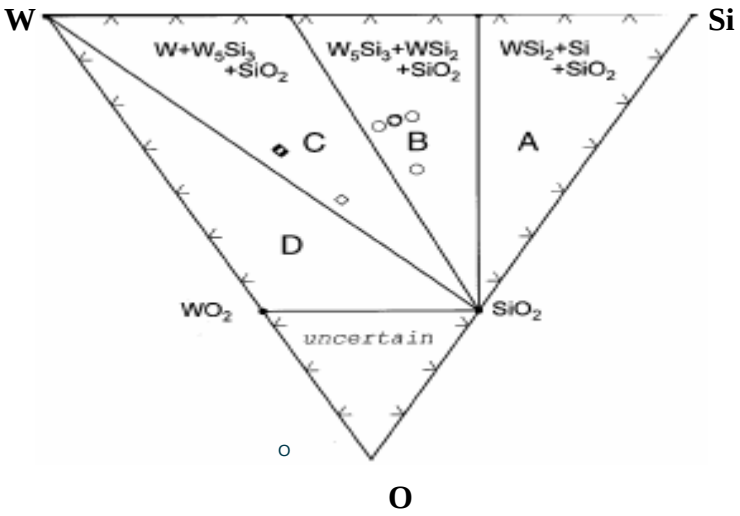


Figure 2 Three element phase diagram. Phase W_5Si_3 can be formed even there is no W rich if WSi_2 is attacked by O_2.

Since there is a nitride layer on top of WSi_2 film (see table 2 process flow), why O_2 get the chance to attack the WSi_2. Figure 3 shows that after gate etching,

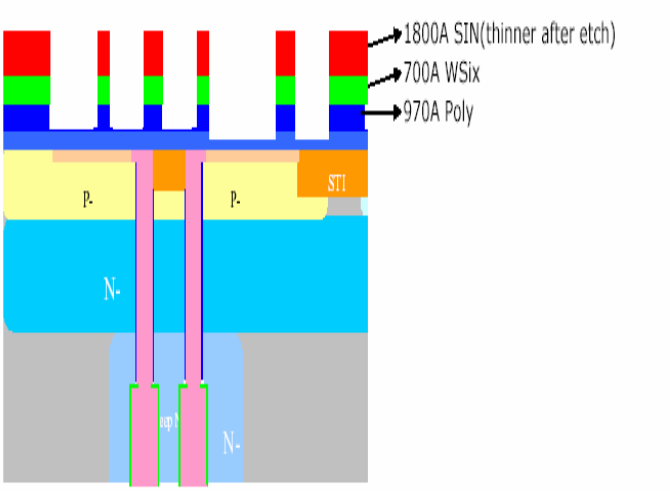


Figure 3 WSi_2 is exposed from sidewall after poly gate etch. Thus O_2 can attack WSi_2 during GC anneal

WSi_2 is exposed from the sidewall. Thus in the subsequent 830 degree GC anneal process, if there is O_2 inside the process chamber, O_2 will react with WSi_2 and form W_5Si_3. WSi_2 will grow abnormally and form the tumor due to phase volume change by bigger lattice constant of W_5Si_3 phase. Thus O_2 control in GC anneal step is the key to eliminate the WSi_2 tumor defects. GC anneal recipe was modified to avoid O_2 sensor delay effect.

Right after the GC anneal, the wafers will grow a sidewall oxide at high temperature in O_2 environment, why WSi_2 will not react with O_2 and forming W_5Si_3 at this stage? Below 710 degree, WSi_2 is an unstable hexagonal phase. An unstable hexagonal phase is

Ex. 1016, Page 317

formed after WSi_2 is deposition at 550 degree. After 830 degree anneal, the unstable hexagonal phase WSi_2 will be transformed to a stable tetragonal WSi_2 phase. O_2 can not attack stable tetragonal WSi_2 even under the high temperature. Figure 4 shows the phase change of WSi_2 at different temperature. At this stage, we fully understand the formation of WSi_2 tumor, and how to eliminate the WSi_2 tumor defects.

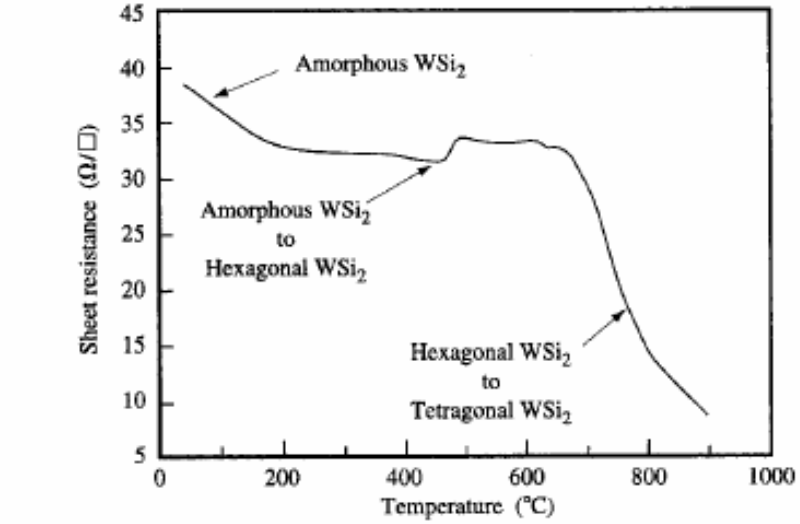


Figure 4 Very stable Tetragonal WSi_2 phase is formed after 830 degree GC anneal

Result and Discussion

By logically analyzing the root cause of WSix tumor formation through study the process flow, GC anneal step was identified as the yield killer step. N2 purge during wafer transfer was set up properly to control the oxygen. Further more, GC anneal recipe was modified to eliminate the affection from the delay of O2 sensor detection. The yield loss of GC CB short due to WSix tumor was reduced from 7% to zero percent.

Acknowledgments

Special thanks to Ruiping Wang, Gene Xu and Michael Huang for their help on the issue trouble shooting.

Ex. 1016, Page 318

ECS Transactions, 18 (1) 291-296 (2009)
10.1149/1.3096465 © The Electrochemical Society

Optimization of Nickel Salicide Strip Process Using Taguchi Method

Shelby Yu, Wenshen Li, Chi-Liang Lin, and Yii-Chian Lu

United Microelectronics Corporation (Singapore Branch), Singapore 519528

A one-step strip scheme of NiPt salicide has been developed to replace an existing two-step strip scheme for better throughput and defects. Taguchi method was utilized to fine-tune three key process factors including thickness of nickel, temperature and duration of the NiSi RTP-1. Optimization was obtained considering the complicated salicide formation such as crystalline structure, dopant type, and dimension of Si. Using signal-to-noise ratio analysis, ANOVA, and multi-objectives compromise, an optimal process was achieved with significant improvements in resistance shift, resistance uniformity, salicide surface roughness, and reverse linewidth effect.

I. Introduction

With the aggressive scale down of MOSFET junction depth, nickel salicide (incorporated with platinum) is finally coming into picture with its low resistivity and silicon consumption (1)(2). Many studies have been published on different NiSi RTP schemes, and it is well documented that 2-step RTP is favorable over a 1-step one (3)(4). However, there aren't as many studies on NiPt strip schemes (2-step or 1-step strip).

TABLE 1. 2-step strip and 1-step schemes.

Process	2-step strip scheme (baseline)	1-step strip schemes
NiPt	d (A)	To be optimized
RTP1	T (ºC); t (sec)	To be optimized
Strip 1	SPM	SPM + HPM
RTP2	T2 (ºC); t2 (sec)	T2 (ºC); t2 (sec)
Strip 2	HPM	skip

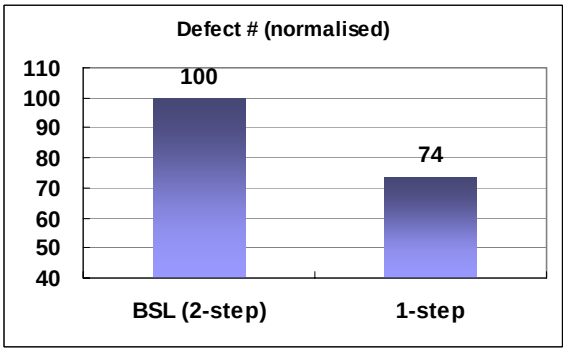


Figure 1. Defect (2-step vs. 1-step strip)

A 2-step strip had been the process baseline of a 65nm logic process on which our experiments were carried out. That is, an SPM (H_2SO_4 + H_2O_2) treatment (strip-1) following RTP-1 to remove un-reacted NiPt, and then an HPM (HCl + H_2O_2) process after RTP-2 to ensure a thorough removal of NiPt residue. A 1-step strip, which put together SPM and HPM in the first strip and skipped the second one, was considered for two merits: First, manufacturability as it cut off one wet process, as shown in Table 1; Second, defect performance, as shown in Figure 1, which might come from an effective removal of NiPt residue and fall-on particles.

However, there were several downside risks of the 1-step strip scheme. Intrinsically, a 1-step strip was not as flexible as a 2-step scheme in terms of salicidation control. With 1-step strip, NiSi Rs (sheet resistance) was found shifted more than 10% from the baseline, uniformity of NiSi Rs was also degraded, and Rs outliers (extremely higher or

Ex. 1016, Page 319

lower) were occasionally found. Besides, reverse linewidth effect became worse, i.e. the difference of NiSi Rs between large and narrow linewidth became larger, and the NiSi surface became rougher, as shown in Figure 2.

To resolve the side effects resulted from a change of strip scheme, three key process factors, NiPt thickness, RTP-1 temperature and duration, could be tuned. For example, Rs goes down with raised RTP-1 temperature, and vice versa. However, to find the process tuning knobs, it was necessary to balance multiple outputs of the complex salicide formation, such as structure of Si (single and poly crystalline), dopant type (N and P), and dimension (four widths from 0.05 to 10um). More than 50 outputs in this case had to be optimized and balanced. Therefore, an effective statistic tool was in great need and Taguchi Method of DOE (design of experiment) seemed an essential solution (5)(6).

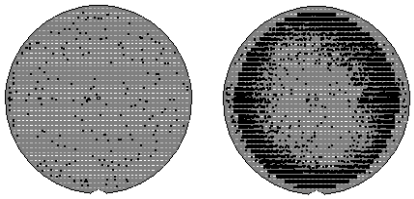


Figure 2. Roughness (scanned by KLA2800)

II. Experiments

Experimental Design (Using Taguchi Method)

To properly evaluate the pros and cons of 1-step strip scheme, 7 outputs (Ys), namely defect, roughness, Rs shift, Rs uniformity, number of Rs outliers, reverse narrow linewidth effect (RNLE) and CP yield, had been addressed. Actually, the total number of outputs was far more than 7, as some of them should be separated into many sub-Ys, due to the various crystalline structures (diffusion and poly), dopant types (N and P), and dimensions (10um, 2um, on-rule and sub-rule) of Si in salicidation, as shown in Table 2.

Baseline (2-step strip scheme) process condition of Ni thickness (A) and RTP1 temperature (B) and duration (C) were fixed at d angstroms, T ºC and t seconds respectively. As for 1-step strip scheme, 3 levels were set with each parameter, as show in Table 3 and total combinations reached 27 (3^3). Using L_9 orthogonal array, the number of splits was reduced to 9, namely A1B1C1, A1B2C2, A1B3C3, A2B1C2, A2B2C3, A3B1C3, A3B2C1 and A3B3C2, where the letters stood for factors and numbers stood for levels. For example, A1B2C2 meant that the conditions were (d-10) angstrom, (T+30) ºC and (t-10) sec for Ni thickness, RTP-1 temperature, and RTP-1 duration respectively.

Experiment Conduction

Splits were carried out at NiPt sputter (thickness) and RTP-1 (temperature and duration) for 1-step strip scheme. Wafers of 2-step strip scheme, serving as reference, were processed in the same lot with wafers receiving 1-step scheme. All test keys of NiSi Rs were fabricated on p-type Si (100) using 65nm CMOS technology. Before NiSi formation, diffusions and poly of both types (N and P) had been readily made after a series of standard processes of STI patterning, Poly deposition and patterning, and doping. (Boron, Arsenic, and Phosphorus implant).

Ex. 1016, Page 320

For baseline reference wafers, NiPt films were deposited by sputter process after being dipped into HF to remove the native oxide. Then TiN cap layers were sputtered on top of the NiPt films without exposing to the atmosphere. Subsequent salicidation was performed by RTP-1. Following RTP-1 came an SPM process (strip 1), in which the TiN and un-reacted Ni was selectively removed. After that, the wafers were subjected to RTP-2, followed by strip-2, in which HPM was used to ensure Ni and Pt residue free at the surface. While in the 1-step strip scheme, the processes remained the same with exception that HPM was added in strip 1 right after SPM, and that strip 2 was skipped.

TABLE 2. Experimental design, Ys

Ys		Sub-Ys	Type of Problem
Y1	Defect count	-	smaller-the-better
Y2	Roughness	-	smaller-the-better
Y3	Rs shift, %	4 type * 4 width	nominal-the-best
Y4	Rs stdev, %	4 type * 4 width	smaller-the-better
Y5	No. of Rs outliers	4 type * 4 width	smaller-the-better
Y6	RNLE, %	4 type	smaller-the-better
Y7	CP yield	-	larger-the-better

TABLE 3. Experimental design, Xs and levels

Parameter (Xs)	Level	Condition
Ni Thickness (A)	Level 1 (A1)	d -10A
	Level 2 (A2)	d -5A
	Level 3 (A3)	d A
RTP-1 Temp (B)	Level 1 (B1)	T + 20C
	Level 2 (B2)	T + 30C
	Level 3 (B3)	T + 40C
RTP Duration (C)	Level 1 (C1)	t - 20sec
	Level 2 (C2)	t - 10sec
	Level 3 (C3)	t sec

III. Results and Discussions

A statistical tool, STATISTICA (7), was applied to analyze the experimental results. Two consecutive analytical steps were involved. Firstly, a single-objective optimization was implemented, during which each output (Y or sub-Y) was analyzed separately and the optimum condition was then selected in terms of the output. At this first step, an early filtration of concerned outputs was essential to reduce the time and effort of analysis. At a subsequent second step, as the optimum conditions might differ, a multi-objectives optimization was carried out to strike a balance.

Single-objective Optimization

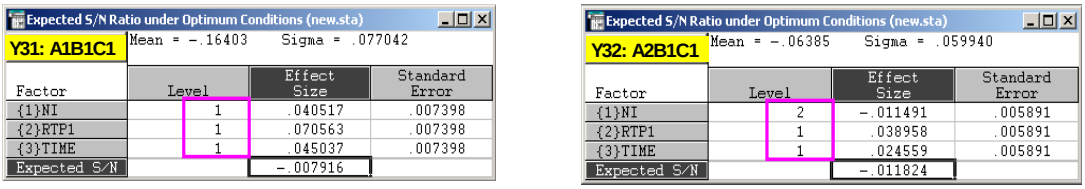
Y1, Defect (Smaller-the-better). It was observed that all 9 splits had better defect performance (33 - 48% defect reduction compared to baseline), and no significant difference was found among the 9 splits.

Y2, Roughness and Y5, Rs outliers (Smaller-the-better). Both issues occurred occasionally in 1-step strip while never happened in the 2-step scheme. It was suspected to be caused by the higher Ni2Si loss along with 1-step strip scheme, which (SPM + HPM) was more aggressive than baseline (SPM only). However, neither issue was found in all 9 DOE splits (with 20~40ºC higher RTP-1 temperature), indicating a higher temperature worked to eliminate both concerns as it enhanced salicidation before strip.

Y3, Rs shift from baseline (Nominal-the-best). It was clear that Rs in small dimension tended to shift more than that in large dimension. Besides, the same behavior was observed in two kinds of small dimensions (on-rule and sub-rule). Hence, special attention should be paid to Rs in small dimension (either on-rule or sub-rule). The Rs of on-rule dimension was chosen for analysis. By doing so, the number of concerned sub-Ys

Ex. 1016, Page 321

was reduced from 16 to 4, which saved much time and effort. The response (expected output) calculation showed optimum conditions, under which the value of sub-Y should be closest to 0, were A1B1C1, A2B1C1, A1B1C1 and A2B2C1 for N+ diffusion (Y31), N+ poly (Y32), P+ diffusion (Y33) and P+ poly (Y34), as shown in Figure 3.



Expected S/N Ratio under Optimum Conditions (new.sta)

Y31: A1B1C1 Mean = -.16403 Sigma = .077042

Factor	Level	Effect Size	Standard Error
{1}NI	1	.040517	.007398
{2}RTP1	1	.070563	.007398
{3}TIME	1	.045037	.007398
Expected S/N		-.007916	

(a). Y31

Expected S/N Ratio under Optimum Conditions (new.sta)

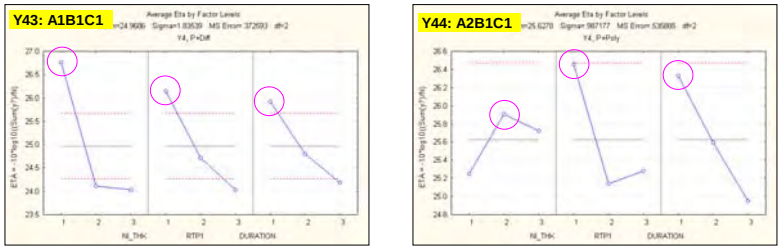
Y32: A2B1C1 Mean = -.06385 Sigma = .059940

Factor	Level	Effect Size	Standard Error
{1}NI	2	-.011491	.005891
{2}RTP1	1	.038958	.005891
{3}TIME	1	.024559	.005891
Expected S/N		-.011824	

(b). Y32

Figure 3 Response analysis of Y3, Rs shift

Y4, Rs uniformity (Smaller-the-better). Similar to Y3, small dimension mattered more than large dimension in Rs uniformity. Likewise, sub Ys were filtered before further analysis. S/N (signal-to-noise ratio) analysis concluded that the optimum conditions were A2B3C3, A3B2C2, A1B1C1 and A2B1C1 for N+ diffusion (Y41), N+ poly (Y42), P+ diffusion (Y43) and P+ poly (Y44), as shown in Figure 4.



(a). Y43 (b). Y44

Figure 4 S/N analysis of Y4, Rs stdev%

Y6, RNLE% (Smaller-the-better). A quantifiable index had been defined as the Rs difference (in percentage) between the smallest and largest widths as shown in Eq. 1. Via S/N ratio analysis, the same optimum condition (A1B1C1) was concluded for each sub-Y, namely, N+ diffusion (Y61), N+ poly (Y62), P+ diffusion (Y63) and P+ poly (Y64).

$$\text{RNLE\%} = (\text{Rs\_Wmax} - \text{Rs\_Wmin}) / \text{Rs\_max width} \quad [1]$$

Y7, CP yield (Larger-the-better). Based on S/N ratio analysis, A2B1C1 was expected to be the optimum condition for CP yield.

Multi-objectives Optimization

As the optimum conditions from various outputs differed, it was essential to strike a balance among all those outputs. Y1, Y2 and Y5 could be put aside since they were not sensitive to the process change. And for Y6 and Y7, A1B1C1 and A2B1C1 was the only optimum condition respectively. For Y3 and Y4, which were relatively complicated, A2B1C1 and A1B1C1 looked the optimum respectively. The reasons were as followed:

1. Y3: Among all 4 optimum conditions derived from 4 sub-Ys, A2B1C1 was, based on response analysis, the most balanced one with less than 5% shift in Rs of all types, as shown in Figure 5.

Ex. 1016, Page 322

2. Y4: By ANOVA, factor A (Ni thickness) was a significant factor for Y43. Hence, it was better to keep level 1 for fact A, which meant A1B1C1 should be preferred as it was the only one with A1. Besides, an analysis of response also concluded that A1B1C1 was the most balanced one, as shown in Figure 6.

Therefore, as shown in Table 4, two candidates remained for the final optimum, namely, A2B1C1, which had the best Rs shift (Y3) and CP yield (Y7), and A1B1C1, which had the best Rs uniformity (Y4) and narrow line width effect (Y6).

TABLE 4. Analytical Process Summary

Ys		Single-objective Optimization	Multi-objectives Optimization	
Y3	Y31	A1B1C1	A2B1C1	**A2B1C1**
	Y32	A2B1C1		
	Y33	A1B1C2		
	Y34	A2B2C1		
Y4	Y41	A2B3C3	A1B1C1 (A2B1C1)	
	Y42	A3B2C2		
	Y43	A*1B1C1		
	Y44	A2B1C1		
Y6	Y61	A1B*1C1	A1B1C1 (A2B1C1)	
	Y62	A1B1C1		
	Y63	A1B*1C1		
	Y64	A1B1C1		
Y7	-	A2B1C1	A2B1C1	

Note: * indicates significant factor

Finally, as A2B1C1 had much better Y3 over A1B1C1 and was quite close to the latter in light of Y4 and Y6, as shown in Fig 5. and Fig 6., it boiled down to that A2B1C1 was the most balanced one. As A2B1C1 was not covered in the original L9 DOE, Si data was still in need as verification.

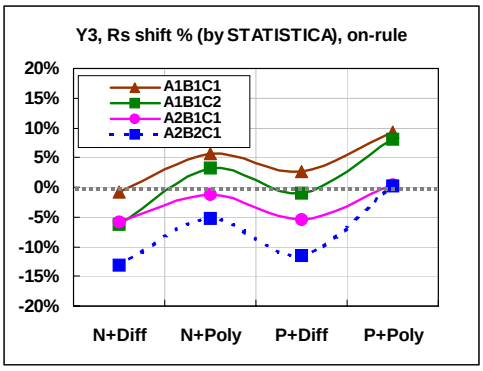


Figure 5 Y3, Rs shift, %

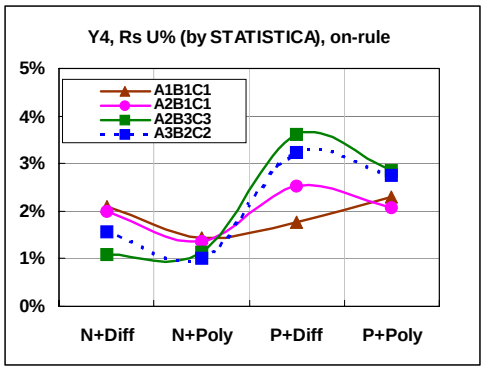


Figure 6 Y4, Rs uniformity%

Confirmation

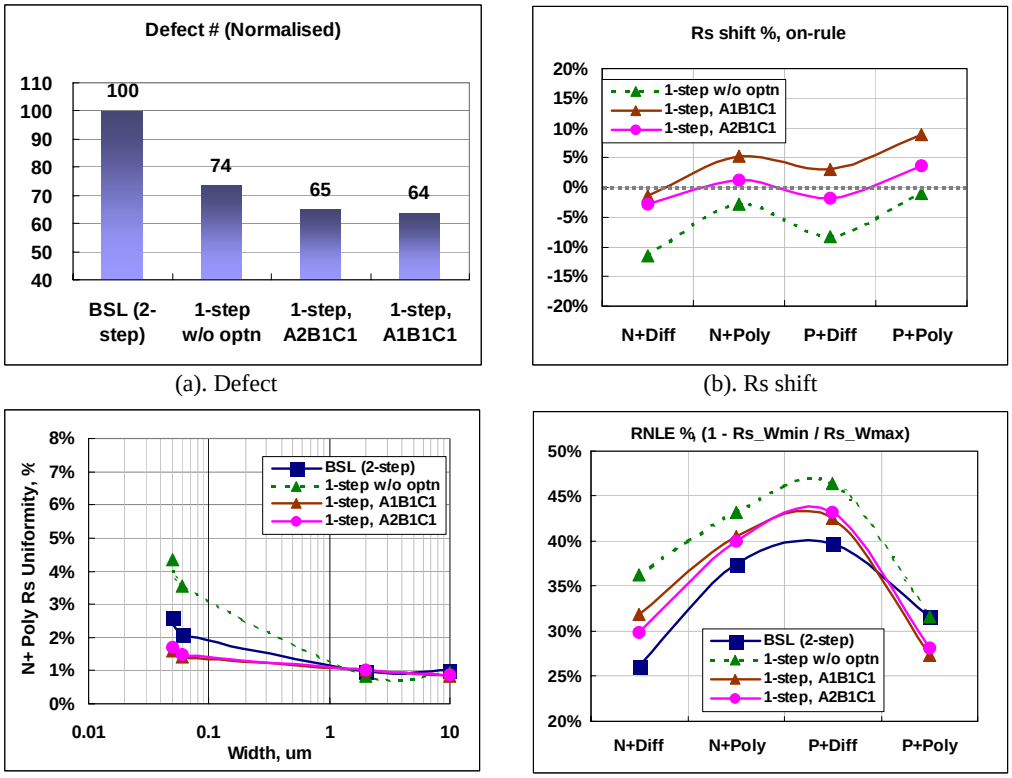
Confirmation experiments verified that A2B1C1 (d-5A, T+20ºC and t-20sec) was the optimum, which had significantly improved defect, Rs shift, uniformity, and narrow width effect, as shown in Fig 9. The results confirmed higher efficiency of 1-step strip in removing NiPt residue and fall-on particles than 2-step one. Regarding the side effects of 1-step strip, it was found higher temperature was imperative to eliminate surface roughness and reduce Rs outliers as it enhanced salicidation before strip. Besides, it was also found properly thinner NiPt thickness and shorter RTP-1 duration was necessary to control Rs shift and linewidth effect.

IV. Conclusions

In summary, a 1-step NiSi strip scheme has been developed to replace a baseline (2-step) for better throughput and defects. Taguchi method has been successfully utilized to

Ex. 1016, Page 323

find an optimal process of Nickel thickness, RTP-1 temperature and duration with significant improvements.



(a). Defect (b). Rs shift

(c). Rs uniformity (d). Reverse linewidth effect

Figure 7, Verification of optimum (A2B1C1)

Acknowledgments

The authors would like to thank R J Guo and Hong Ma (UMC Singapore) for their guidance to experimental design and data analysis, and to thank Jianmin Ji, Haifeng Shi and Tony HT Liu as well for their efforts in process tuning and wafer processing.

References

1. H. Iwai et al., *NiSi silicide technology for scaled CMOS*; Microelectronic Engineering, 60 (2002).

L.J. Chen, *Silicide Techonology for Integrated Circuits* (2004).

2. K. Funk et al., NiSi Contack Formation – Process Integration Advantages with Partial Ni Conversion. IEEE (2004).
3. J.A. Kittl et al, *Phase effects and short gate length device implementation of Ni fully silicided (FUSI) gates*, IMEC, Kapeldreef 75, B-3001 Leuven, Belgium (2006).
4. K. N. Otto et al., Extensions to the Taguchi Method of Product Design, Engineering Design Research Laboratory, California Institute of Technology (1991).
5. C. T. Su, *Quality Engineering*, Chinese Society for Quality (2001).
6. Stat Soft, http://www.statsoft.com/products/products.htm

CHAPTER 4

PHOTOLITHOGRAPHY

ECS Transactions, 18 (1) 299-314 (2009)
10.1149/1.3096466 © The Electrochemical Society

Inverse Lithography Technology (ILT) Enabled Source Mask Optimization (SMO)

Linyong Pang, Guangming Xiao, Vikram Tolani, Peter Hu, Tom Cecil, Thuc Dam, Ki-Ho Baik, Bob Gleason

Luminescent Technologies, Inc. Palo Alto, CA 94303, USA

In the first implimentation by Luminescent of ILT-enabled Source-Mask Optimization (SMO), an ILT-optimized mask was generated for each designated illumination condition as the source was swept through various parameter settings in order to find the best combination of source and mask. This approach has been successfully applied to explore and select lithography processes and design rules for advanced semiconductor technology nodes. In Luminescent's latest implimentation of ILT-enabled SMO, the same Level Set Method used in the mask optimization is used in the source optimization; in other words, the source map is represented by a level set function. During the optimization process, the level set function evolves to achieve a minimized cost function, where the cost function is defined as the difference between aerial image and ideal image (MEEF and DOF can also be used in the cost function) via a gradient flow, and the gradient flow is based on the cost function itself. This flow is interlaceded with the mask inversion flow so that a simultaneous mask & source co-optimization is achieved. In this paper a number of memory and logic device results at the 32nm node and below are presented to demonstrate the benefits of ILT-enabled SMO.

Introduction

For semiconductor manufacturers moving to advanced technology nodes –45, 32nm, 22nm and below – lithography presents the greatest challenge, because it is fundamentally constrained by basic principles of optical physics. It has long been recognized that the best lithography that is theoretically possible can be achieved by considering the design of photomasks as an inverse problem -- and then solving the inverse problem to find the optimal photomask and source for a given process, using a rigorous mathematical approach.

For years, source optimization and mask pattern correction have been conducted as two separate RET steps. For source optimization, the source was optimized based on fixed mask patterns; in other words, OPC and SRAFs were not considered in source optimization. This solution results in a source optimized only for dense patterns, but which may not work well through pitch because of the tight coupling between illuminator design and SRAF placement. This is important because defocus behavior of sparser patterns often limits process window.

In the first implimentation by Luminescent of ILTenabled Source-Mask Optimization (SMO), an ILT-optimized mask is generated for each designated illumination condition as the source was swept through various parameter settings, in order to find the best

Ex. 1016, Page 327

combination of source and mask. This approach has been successfully applied to explore and select lithography processes and design rules for advanced semiconductor technology nodes. In Luminescent's newest implimentation of ILT-enabled SMO, the same Level Set Method used in the mask optimization is used in the source optimization; in other words, the source map is represented by a level set function. During the optimization process, the level set function evolves to achieve a minimized cost function, where the cost function is defined as the difference between aerial image and ideal image (MEEF and DOF can also be used in the cost function) via a gradient flow, and the gradient flow is based on the cost function itself. This flow is interlaced with the mask inversion flow so that a simultaneous mask & source co-optimization is achieved. In this paper a number of memory and logic device results at the 32nm node and below are presented to demonstrate the benefits of ILT enabled SMO.

Parameters and Utilities in SMO

The goal of SMO is to improve lithography performance during wafer manufacture. A number of parameters are used to gauge lithography performance, and a good SMO program should be able to optimize any weighted combination of such parameters. Parameters typically used to assess lithography performance include image contrast, Image Log Slope (ILS), Depth of Focus (DOF), Exposure Latitude (EL), Mask Error Enhancement Factor (MEEF), and Process Variation (PV) bands.

In particularly, as low k1 lithography resolution approaches the physical imaging limit, image contrast is poor and MEEF becomes intolerable, so it becomes increasingly important to consider MEEF and mask error in the total budget of process variations. Luminescent's SMO program provides two means of quantifying and evaluating the impact of mask error, called Pixed-based MEEF Map and "MEEF-aware" PV Bands.

Edge-based MEEF

MEEF is defined as the wafer CD variation as a function of the mask CD variation to wafer level (1X). MEEF can be expressed mathematically as [11, 12]:

$$MEEF = \frac{\partial CD_{wafer}}{\partial CD_{mask}} \qquad [1]$$

Although we define MEEF with respect to a global mask bias, points along the perimeters of image patterns can move by different amounts. A CD measurement is the difference between two edge positions, which may move by different amounts under the influence of mask bias, complicating the interpretation of MEEF. Furthermore, we are often interested only in the position of a single edge, such as when characterizing line-end pull-back. In order to capture the sensitivity of each edge to mask error, an edge-based MEEF calculation is proposed. It can be written as

$$MEEF_{edge} = \left. \frac{\partial EPE_{wafer}}{\partial EPE_{mask}} \right|_{normal} \qquad [2]$$

where EPE is used instead of CD. ∂EPE_{mask} is the mask perturbation per edge, equal to ½ ∂CD_{mask} in the traditional definition, i.e. the mask is evenly biased. ∂EPE_{wafer} is measured as the distance between the image contour of a perturbed mask and the image contour of a nominal mask along its normal direction, as illustrated in Figure 1.

Ex. 1016, Page 328

Therefore, the edge-based MEEF can be extracted from the ratio of perturbation from wafer to mask. Along the edge of nominal, two edge-based MEEF numbers can be calculated for positively (+) and negatively (-) biased masks, respectively. The two MEEF numbers are usually slightly different from each other due to asymmetric response to bias-up and bias-down. The magnitude of difference in MEEF of ± biased mask depends on local environment and is larger for high MEEF area. Since mask error can be either positive or negative, central difference approximations to the derivative are appropriate.

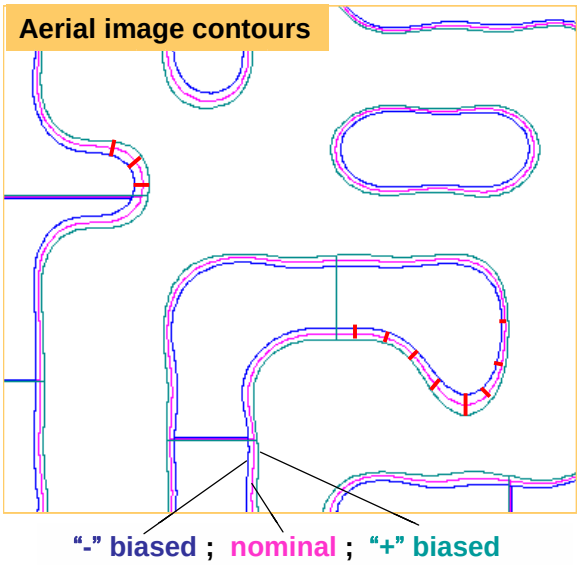


Figure 1. An example of aerial image contours from regular and biased masks. EPE based MEEF is calculated from the delta between positively (+) and negatively (-) biased masks along normal direction of nominal image contour.

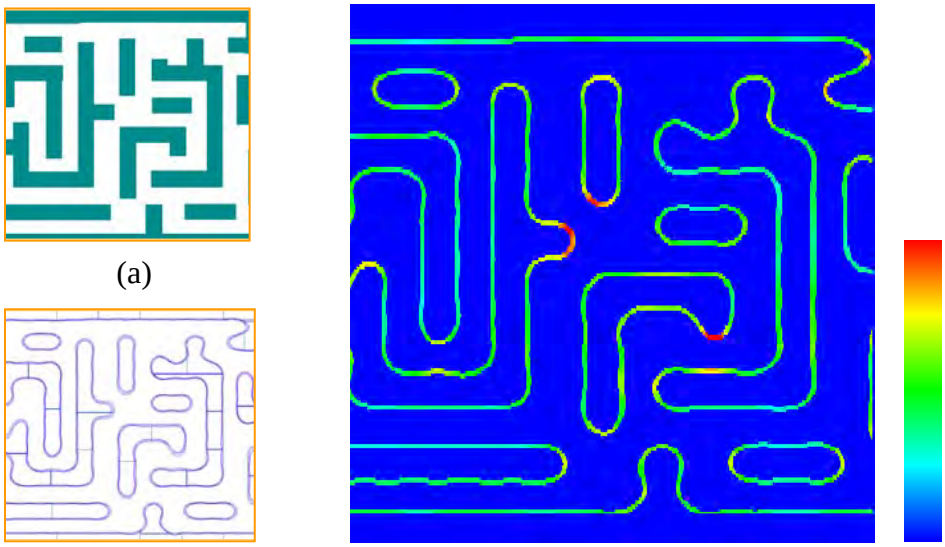
Full Field Pixel-based MEEF Map

Traditionally, MEEF is calculated by measuring CD change across a cut-line. Multiple cut-lines are placed across potentially challenging areas. This method is often sufficient for relatively simple pattern. For a complex two dimensional pattern, however, due to the strong proximity effect among neighboring features and high inter-correlation among illumination conditions, mask, and OPC at low-k_1, it becomes more challenging to pre-identify all of the weak points. Because the number of cut-lines is practically limited, the cut-line approach may not measure all critical locations. A full-field rather than a limited cut-line analysis of MEEF is therefore important to assure the capture of all critical information needed for a complex 2-dimensional design.

From the definition of edge-based MEEF given in the previous section, one can derive the MEEF of every segment along an image contour. Due to the variation of MEEF, the image contours must be divided into small segments in order to accurately calculate MEEF everywhere. The native implementation of ILT in pixel-based space makes it ideal to render edge-based MEEF into a full field MEEF bitmap. As shown in Figure 2, a full field color MEEF map generated from Luminizer™ Inverse Explorer displays the color coded MEEF value of each pixel along the aerial image contour. MEEF hotspots are visually apparent on the map. In the previous section, we mentioned asymmetry of MEEF across a pair of edges due to the different proximity environment.

Ex. 1016, Page 329

It is evident from the MEEF map that at these highly non-uniform locations, two sides of a contour do not display the same colors. Traditional CD cut-line based MEEF will average out the two sides of a contour, underestimating MEEF. This can potentially hide serious design problems.

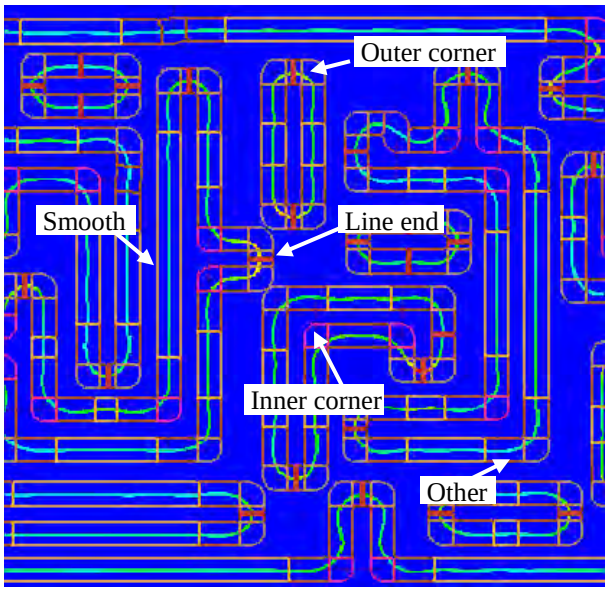


(a) (b) (c)

Figure 2. (a) Design target clip; (b) Aerial image contours of +/- biased masks; (c) full field edge-based MEEF bitmap showing MEEF distribution and hotspots. Red color represents high MEEF and cool color represents low MEEF value. Each pixel along the contour represents the edge-based MEEF of the edge segment inside the pixel.

The MEEF map is visually useful for identifying hotspots. In addition to the MEEF map, Luminizer™ Inverse Explorer also generates MEEF statistics. The tolerance for process variation can be different for different parts of a design. For example, the tolerance for the inner corner of a metal line can be much higher than for line-end. To provide a better assessment of MEEF, automatically categorized MEEF reports are generated for each type of topology, such as line-ends, outer corners, inner corners, etc. Figure 3 shows different types of geometry-based topology markers and reported MEEF statistics for each topology, with maximum and average MEEF values. Aside from automatically generated topology markers, customized user defined markers can also be added to the design for further customized analysis. Additional MEEF reports can be generated for each type of user defined topology regions.

Ex. 1016, Page 330



Topology MEEF Statistics:

	Maximum	Average
Line end:	6.80	3.59
Smooth :	4.82	2.38
Outer corner:	5.71	3.39
Inner corner:	4.22	2.88
Others:	6.79	2.75

Overall MEEF Statistics:

Maximum MEEF : 6.80
Average MEEF : 2.66
Standard deviation : 0.71

Figure 3. Automatically generated geometry based topology markers overlay with MEEF map. MEEF statistics of each topology type as well as overall combined of all topology types are generated for detailed analysis.

MEEF Aware Process Window and PV Band

Process window is the most important metric in lithography simulations for evaluating the performance of a given RET solution. Achieving a certain minimum depth of focus (DOF) at a given exposure latitude (EL) is the paramount goal. Process window calculation traditionally is based on a perfect mask, assuming no mask errors, no corner rounding, etc. In reality, process variation always leaves residual mask errors which affect process window for on-wafer printing. When MEEF is low, the impact of mask error is small compared to the total budget of process variation. In a low k_1 regime, MEEF increases substantially to the level of 4~6, or even higher. Mask errors are magnified several-fold and have a much more pronounced impact at wafer level. DOF predictions that do not consider MEEF are usually unrealistically optimistic.

To obtain a more realistic view of whole process variations, mask error must be factored into process window calculations. Given mask bias, *m*, inferred from mask targeting specifications, MEEF aware process window is calculated as the common process window of unbiased and $\pm m$ biased masks. As shown in Figure 4, when MEEF and a 1nm mask bias were included in the computation, DOF at 5%EL was reduced from ~220nm (Figure 4(a)) to ~145nm (Figure 4(c)).

In addition to process window, another useful way to study the predicted quality of printed image is by analyzing PV bands. The PV band is generated by calculating wafer aerial image at the various process conditions and combining the resulting images into a band. The width of the band represents the range within which a feature will print as the process conditions vary. Typically the conditions used to generate PV bands are %EL and defocus, as shown in Figure 4(b). If MEEF and mask error are also included in the process variation conditions, a much larger band area is generated as seen in Figure 4(d), which suggests a much worse image variation than would result from a perfect-mask assumption. Figure 4(b) and 4(d) represent the best and worst cases from mask variations,

Ex. 1016, Page 331

respectively. Actual variation would be expected to fall between the two extreme scenarios.

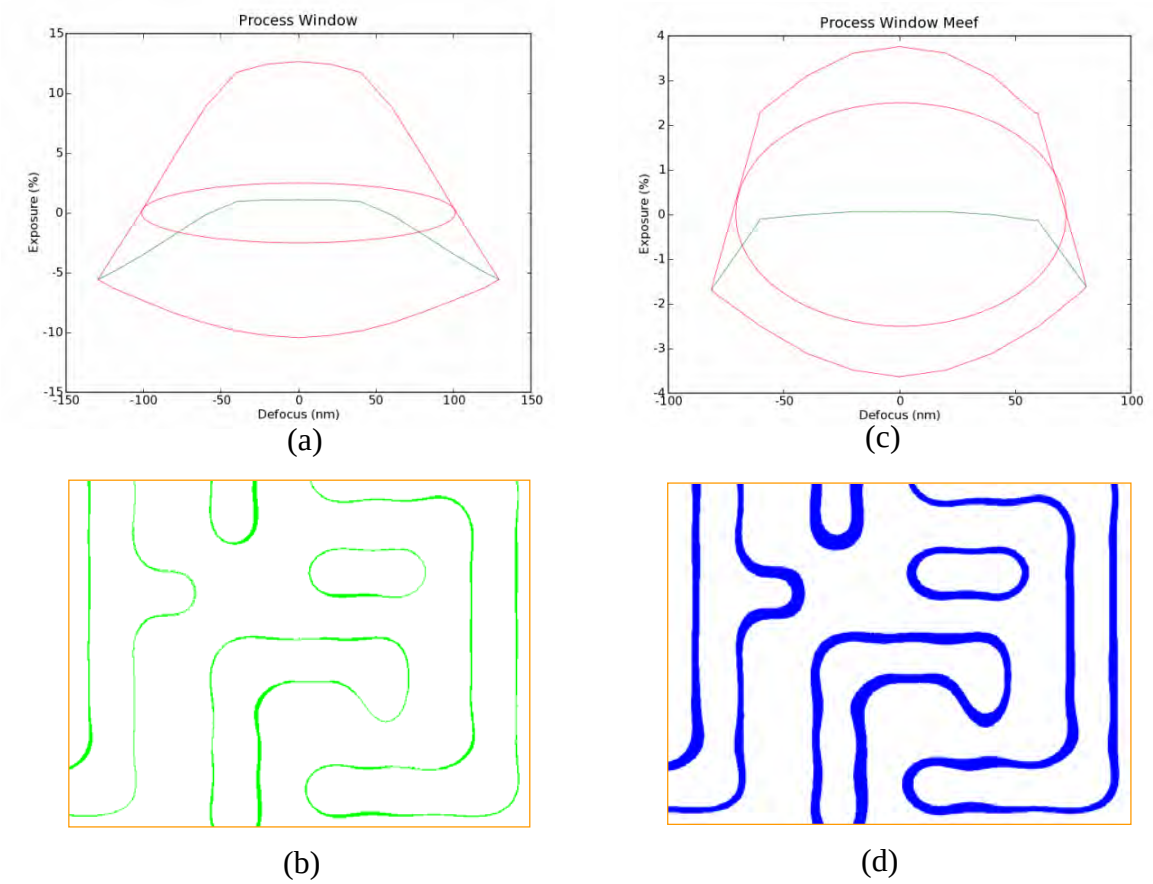


Figure 4. Process window and PV band with and without MEEF term included. (a) Regular process window plot; (b) PV bands with no mask error in account; (c) Common process window plot with awareness of mask error. It is a common process window of three masks, unbiased and +/- biased. (d) PV bands with mask errors included in process variation conditions.

Luminescent SMO Framework

In SMO, both source and mask must be optimized simultaneously. In Luminescent's SMO approach, the mask optimization is performed using the patented Level-Set-Methods-based Inverse Lithography Technology (ILT), and both the mask and source optimization rely on the Level Set Method. Both optimizations are interlaced and run iteratively to achieve the best lithography performance.

Overview of ILT using Level Set Methods

Luminescent's approach to ILT is based on a branch of mathematics co-invented by Stan Osher at UCLA. Commonly known as Level-Set Methods [13], these techniques have been applied to solve inverse problems in a wide range of engineering disciplines such as image processing and fluid dynamics. In Luminescent's formulation, the design intent target and simulated wafer contour are represented in 2D pixilated arrays (2D gray scale image). The post-correction mask is represented as a 2D level-set function, and

Ex. 1016, Page 332

mask transmission as a 2D, potentially complex, array (Figure 5). This formulation of the problem has a variety of advantageous properties. For example, the level-set representation allows for contours to merge, break, appear, or disappear, in a consistent, mathematical representation. Various functions (for example, the wafer image) can be determined as closed-form expressions. The mask function itself is an element of a Hilbert-space which is much larger than the two-dimensional space of the photomask, allowing for "more global" solutions to be found [14].

A lithography-forward model, which simulates the mask effects, scanner optics, resist development, and even etching and SEM bias, are created and used in ILT. Such a forward model is similar to models used in OPC and lithography simulators. Once a mask pattern is given, a forward simulation can be run to compute the aerial and latent image which in turn can be used to predict a printed resist image, and even an etched image.

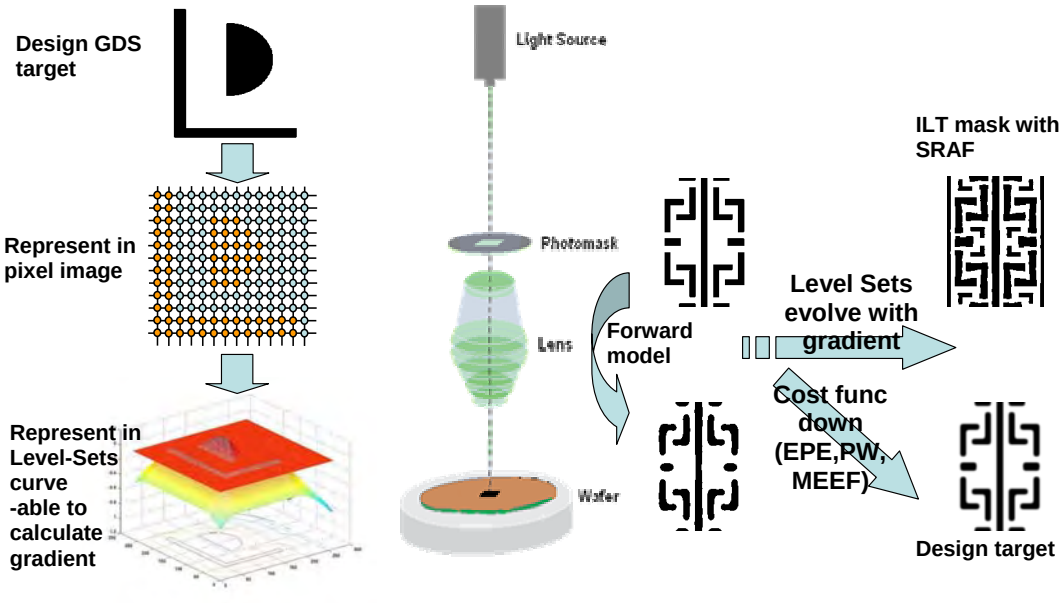


Figure 5. The level-set representation of a mask, and the flow of using Level Set Methods to solve the lithography inverse problem.

A cost function (also called a merit function, energy function, or Hamiltonian) can be defined which provides a measure of the quality of the solution, or the "goodness" of the mask. In a simple case, the cost function could be the absolute value of the difference between the wafer image and the target pattern, integrated over the area of the mask. In practice, a number of additional elements may be included in the cost function. For example, the wafer pattern at various conditions throughout the process window (i.e., over or under exposed and/or plus/minus focus), the Normalized Image Log Slope (NILS) of the image, the robustness against MEEF, or other factors as deemed appropriate.

ILT can use multiple simulated wafer images at different process conditions in the cost function (E). As shown in Figure 6, simulated wafer images at nominal, defocus, off-

Ex. 1016, Page 333

nominal exposures, and even the image simulated with mask biases (e.g., +1nm and -1nm)can be used in the cost function to ensure the inversion giving solutions that print well under those process variations.

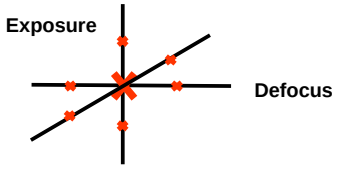


Figure 6. Multiple image planes used in ILT, including images planes at nominal, defocus, off nominal exposures, and images simulated with biased masks.

$$E = \sum_{x,y} \sum_{E,D,M}^{n} w_{(Ei,Di,Mi)} \sum_{topo=1}^{m} w_{topo} \left(f(\text{mask}) - \text{target} \right) \qquad [3]$$

The actual functional form may be different from the form as described above as well. Elements that are not directly related to lithography may be included; for example, simple masks may be preferred over complex masks, and terms to this effect may be included in the cost function as well. What is essential is that the cost function is a function of the mask function, and minimizing the cost function allows us to find the optimal mask according to the criteria we have chosen. Representing the mask pattern, wafer contours and target as level sets allows us to compute the gradient of the cost function with respect to the mask patterns. This enables us to use modern minimization algorithms to solve the lithography inverse problem in a fast and efficient way, and apply the results to full chip correction.

Constraints represent another important aspect of the minimization. A variety of constraints are imposed by the realities of mask manufacturing; for example, two disjoint chrome regions must be separated by a minimum distance, and a chrome line must have a minimum width. These constraints can be addressed by defining a sub-space of the full Hilbert space of mask functions, and restricting the solution to this sub-space.

A key distinctive feature of ILT is the absence of pattern-dependent heuristics, and the ability to broadly explore wide areas of solution space. This leads to a very important advantage of ILT in the application of Sub-Resolution Assist Features (SRAFs), mask features that do not print on the wafer and which are detached from the edges of the main mask patterns, and yet manipulate the light reaching the wafer so as to accentuate the wafer image. In the past, these were placed empirically, with great care, and frozen in place during the computation of the rest of the mask. In contrast, ILT can determine optimal SRAFs simultaneously with the rest of the mask. Level Set Methods naturally support SRAFs in the inversion calculation. The Level Set surface evolves during the inversion calculation, which allows the SRAFs to generate, change dimensions, and even vanish. For example, when the mask pattern level set surface crosses the “0” level set, which defines the mask contour, SRAFs appears; when the surface is changing, SRAFs change shape and size; and when the surface does not cross the “0” level set, SRAFs disappear. In ILT, SRAFs are treated the same as main features, so SRAFs and main features are co-optimized during inversion. In addition, SRAFs are consistent with the gradient flow used for the inversion calculation, so the SRAF signal can be naturally carried by the gradient of the cost function. The absence of segmentation scripts is

Ex. 1016, Page 334

another significant advantage because it usually requires significant engineering resources to write such scripts for different patterns on different design layers[15-20].

Case Study of ILT with and without MEEF in Cost Function

The following example shows the impact of using MEEF in ILT. In this study, design clips from an advanced memory technology 4Xnm node were used. Maximum MEEF and DOF values for all topologies and line-end-only are shown in the charts (Figure 7). Results of running ILT with and without MEEF in the cost function are compared for various source conditions, such as C-Quad, QUSAR™ and dipole. By looking at the overall distribution of the MEEF and DOF in the ILT solution, one can observe a reduction of about 25% for maximum MEEF and MEEF for line-ends when MEEF was included in the cost function, while the common DOF was reduced ~10% with the corresponding MEEF reduction. This illustrates the tradeoff often required between MEEF and DOF in any litho optimization for advanced technology nodes. Nevertheless, it is possible to find certain source conditions under which the MEEF is reduced while DOF is improved (looking at individual dots in the chart).

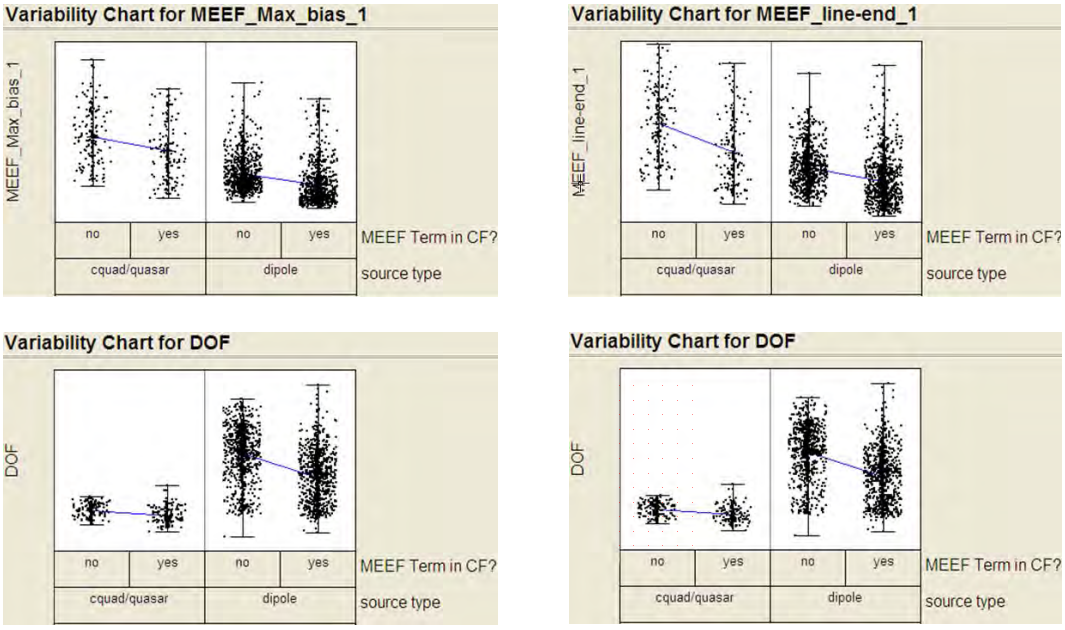


Figure 7. Impact of the inclusion of MEEF in the ILT cost function on MEEF and DOF for advanced memory design clips under various source conditions.

Source Optimization and ILT-Enabled SMO

If ILT is used only for mask optimization, the improvement may be limited by the source, because some sources (e.g., disk) do not benefit from SRAFs, while other sources (e.g., QUASAR™) require more complicated SRAFs than others (such as annular). In such cases, SMO can maximize the improvement possible from ILT while striking the right balance between mask and source complexity.

In Luminescent's implementation of SMO, the Level Set Method is used in source optimization – in addition to represent the mask patterns with a level set function, the source map is also represented by level set function. During the minimization process,

Ex. 1016, Page 335

the level set function evolves to achieve a minimized cost function (e.g., the difference between aerial image and ideal image) via a gradient flow, and the gradient flow is based on the cost function itself. This flow is combined with the mask inversion flow so that a simultaneous mask & source co-optimization is achieved. In such a flow, both MEEF and DOF can be integrated into the cost function of SMO as nonlinear terms.

SMO Case Studies

SMO Case Study for Advanced 4Xnm Memory – SMO can balance complexity between source and mask

Figure 8 shows an example of SMO for an advanced memory device. The target pattern is a 4Xnm contact array with different pitches in X and Y directions. Three combinations of source and mask are shown in this example. In the first scenario, the illumination is a fairly simple QUASAR™, but the mask pattern is quite complicated, especially the SRAF patterns. A reasonable process window (DOF 141nm @6% EL) was achieved. In the second scenario, the mask pattern, including rectangular SRAFs is fairly simple; however, the source map is very complicated. The process window is similar to that in scenario one. In the third scenario, both mask pattern and source map are fairly simple. Higher mask complexity increases mask write time and makes mask inspection more difficult. This case study demonstrates that one can push the complexity from mask to source in order to reduce OPC and mask-making cost, a very practical benefit in the case of memory arrays[21].

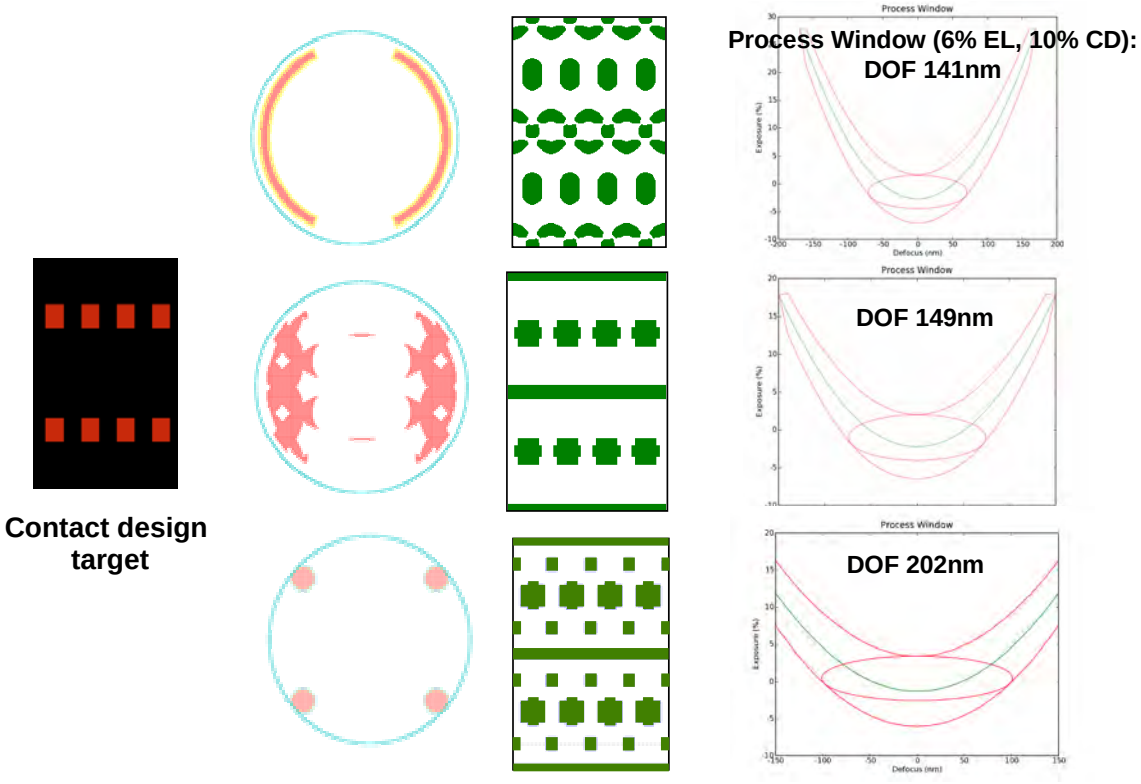


Figure 8. Source & mask co-optimization results for a 4Xnm contact array showing three scenarios.

Ex. 1016, Page 336

SMO Case Study for Advanced 4X Memory – SMO can further improve process window and MEEF than using ILT alone

The following is another SMO case study for an advanced 4X memory device. In this case, ILT and SMO were compared with the customer's POR source and mask. As shown in Figure 9, when the POR source was kept and ILT was applied, DOF increased about 30%, while MEEF was reduced from 7.3 to 5.1. When a free-form SMO was applied, DOF and MEEF were further improved. Free-form sources obtained from SMO generally give the best performance; however, it may be difficult to manufacture such a complicated DOE. In such a case, a DOE-constrained source optimization may provide a more realistic solution. In this example, a parametric source, a rotated QUSAR^{TM} plus disk illumination, is also shown as a solution, and is clearly a better solution than the POR in terms of both DOF and MEEF[22].

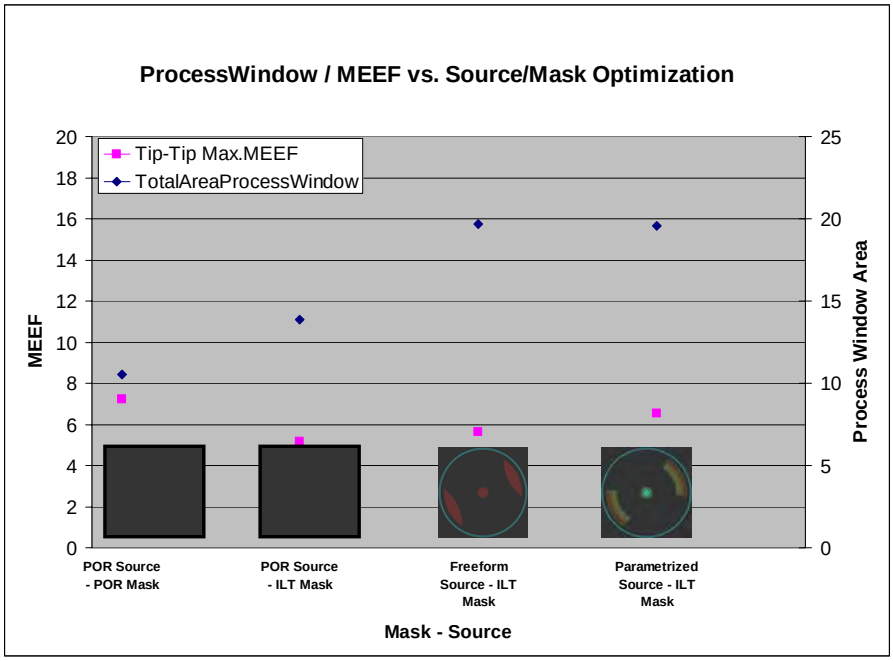


Figure 9 shows a case study of SMO for an advanced memory device. The target pattern is a customer's 4Xnm contact design.

SMO Case Study for Advanced 22nm Logic – A typical SMO Workflow and Impact of Use MEEF in SMO

SMO for a logic device is usually more complicated than for a memory device, because in the case of memory, the target of SMO is usually the memory array, but for logic, SMO must consider many variations of patterns, including 1D through pitch patterns, SRAMs, and some random logic clips. Also, because so many different patterns must be considered for logic SMO, the optimized source for logic SMO tends to be more regular than the optimized source for memory SMO. This might lead to the hasty conclusion that that logic SMO is not as important and useful as memory SMO, but at the 32nm technology node and beyond, lithography optimization can still be very useful. Off-axis illumination has to be used to improve imaging, and design rules have to change to compensate for the restrictions of lithography rules.

Ex. 1016, Page 337

Before doing SMO for a combination of various patterns, SMO for each individual pattern is first presented. Figure 10 shows the design target of three 1D patterns at critical design rule pitches. The most dense line/space pattern has only one direction in order to obtain a better lithography performance. The other looser pitch patterns have both x and y components. For such simple dense 1D patterns, there are theoretical answers to the best source from partially coherent 2-beam imaging or 3-beam imaging theory. As shown in Figure 10, the SMO algorithm obtained answers consistent with the theoretical results.

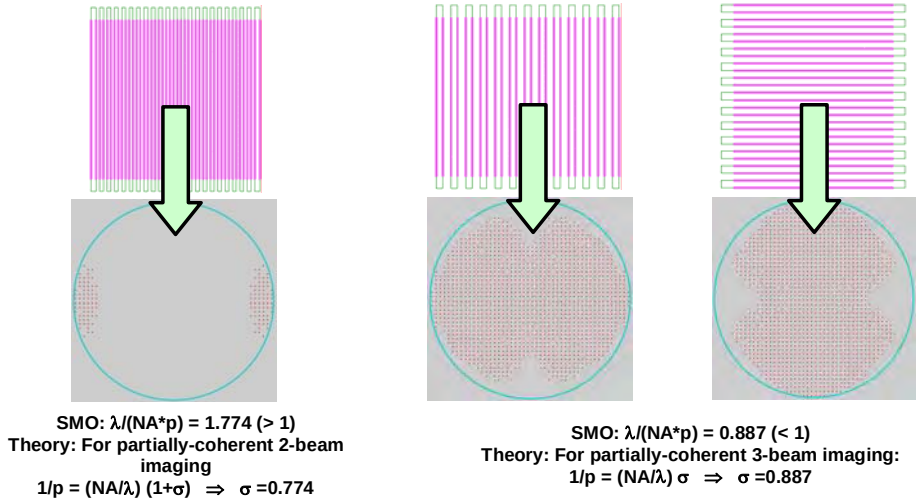


Figure 10. Free-form SMO on 22nm key pitches of 1D patterns

The next step is to review the SMO result for random logic clips. Typically, random logic clips are complicated 2D patterns that include not only dense, but also semi-dense and isolated patterns. In such a case, there is no simple theoretical answer. In addition, source optimization only will not provide the right answer, because mask bias (i.e., OPC) and SRAFs are needed for the semi-dense and isolated patterns to maximize the common process window. ILT, due to its integrated approach to optimize main patterns and SRAFs without any pre-defined rules (except mask rules), becomes the best choice for SMO by combining mask optimization with source optimization. Figure 11 shows the SMO result for a random logic clip. In this case, the smooth region (as marked on the design clip) was used in the cost function. A free-form source and an inverse mask satisfying customer mask rules (minimum CD and space: 20nm, minimum corner to corner: 15nm, minimum segment length: 10nm) were obtained.

Ex. 1016, Page 338

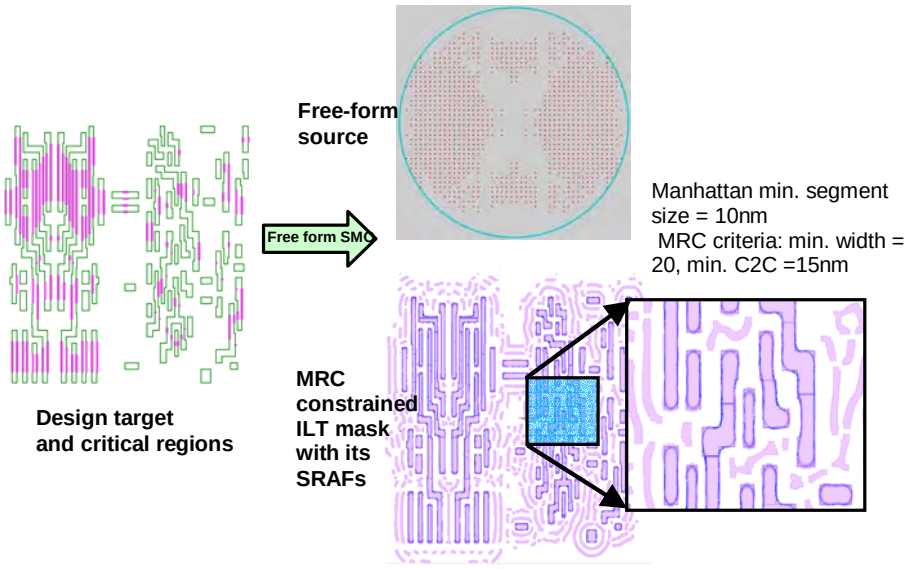


Figure 11. SMO results for a 22nm random logic clip.

The examples above show the SMO results for individual clips. For 22nm logic SMO, variations of critical patterns must be included. In the following, a typical SMO workflow is explained (Figure 12). First, various patterns, including 1D through pitch patterns at critical design rule CDs and pitches, some SRAM patterns, and samples of random logic clips are used to calculate free-source SMO. This free-form source for multiple design clips is noticeably different from any of the free-form sources for each individual pattern. The free-form source gives the best lithography performance, but it may be difficult to manufacture the necessary diffractive optical elements (DOEs). Therefore, one could use a free-form source as a guideline, and approximate it with a parametric source. In this case, a Disar plus QUSAR^{TM} resembles the approximate shape of the free-form source. Now a DOE constrained composite source SMO is performed to find the right values for those variables in the composite source, such as open angle, sigma in and sigma out, to give the best lithography performance. After finding the optimized composite source, a regular source satisfying DOE constraints is obtained.

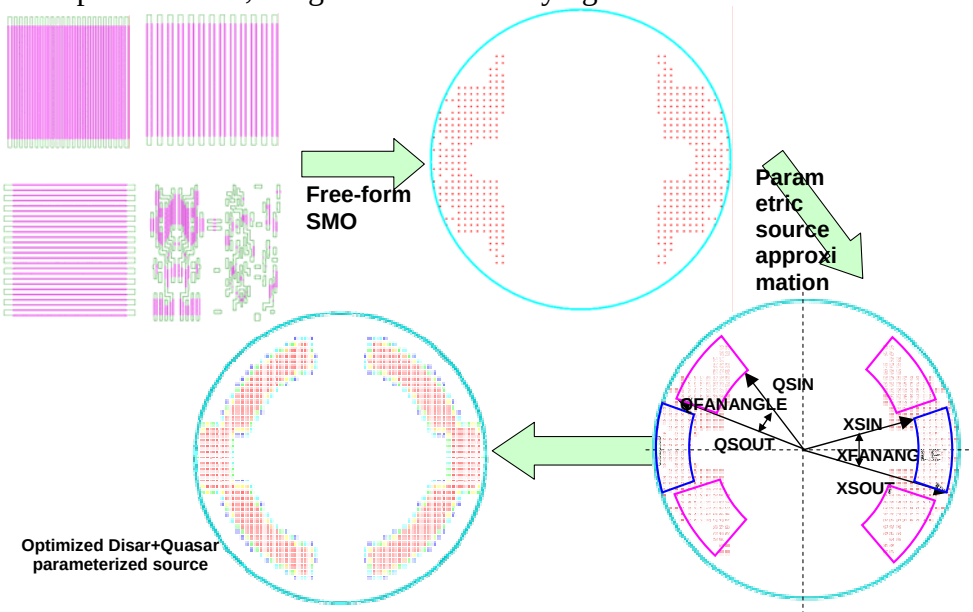


Figure 12. SMO workflow for 22nm logic

Once the SMO flow is established, one can study the impact of MEEF in SMO. Figure 13 shows the SMO results with and without consideration of MEEF. The PV band was used in this study as the parameter to optimize in SMO. Since process margin requirements for focus, dose latitude and mask bias can all be specified, PV band is a good way to understand, visualize, and quantify the process margin. It is also less computationally expensive than calculating DOF, exposure latitude and MEEF; therefore, it is now commonly used in SMO calculations. From Figure 13, one can clearly see, after including the MEEF, that the PV band is significantly reduced.

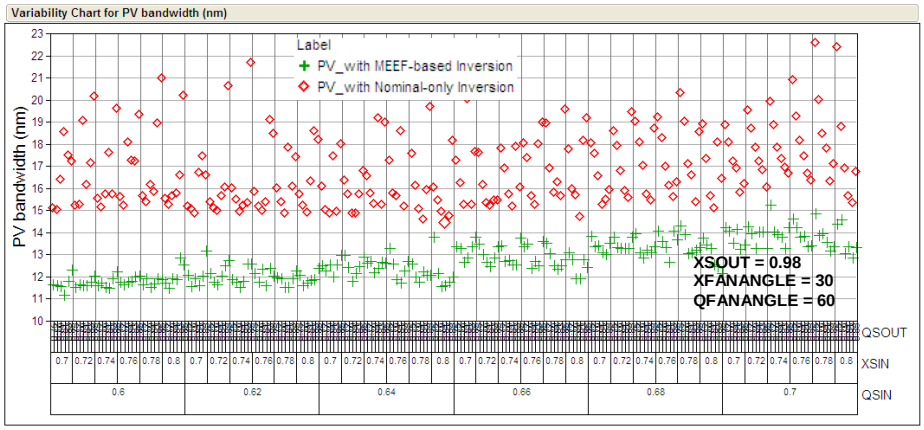


Figure 13. Comparing SMO results for 22nm logic with and without considering MEEF

Figure 14 shows the summary of the final results of SMO compared with customer source and mask for a 22nm logic pattern. Again, the PV band is used as the indicator for SMO result. It is clear that Luminescent's SMO source and mask outperform the customer's source and mask across the board. For reference, SMO results based on other DOE constrained sources, such as annular, C-Quad, Disar, Disar plus annular, are also shown in this chart. It can be seen that the SMO source also outperforms those sources in addition to the customer's source.

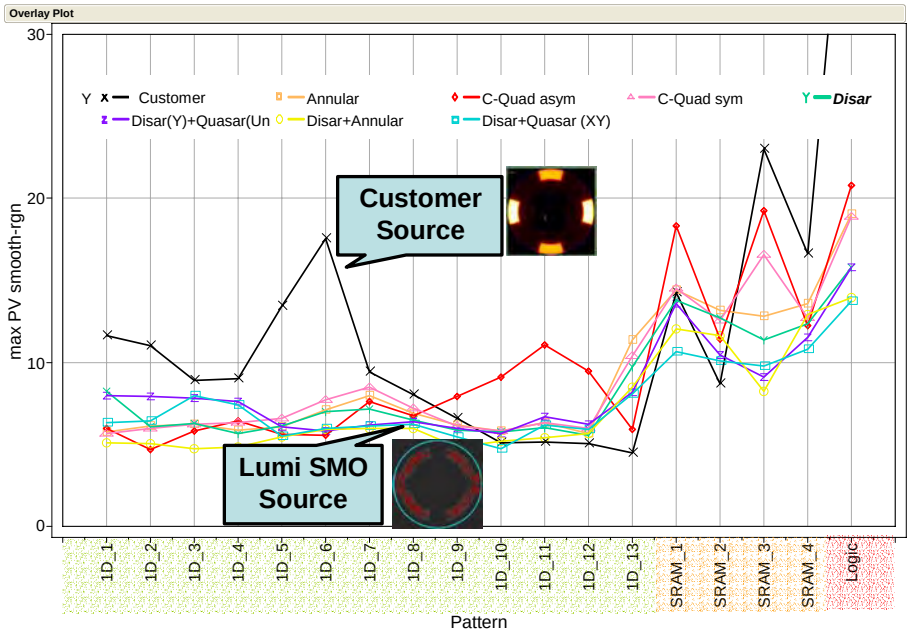


Figure 14. Summary of SMO results for 22nm logic

Summary and Conclusions

Source optimization and mask optimization (e.g., OPC, PSM, ILT) were formerly considered separate resolution enhancement techniques in lithography. For immersion lithography to be extended to 32nm and below, they must be combined to achieve the largest possible enhancements.

ILT is a key enabling technology in Source and Mask Optimization, because it eliminates the trial-and-error steps in mask optimization, and provides an integrated solution for OPC and model-based SRAF.

In Luminescent's approach, the same Level Set Methods used in ILT to represent mask shapes, are applied in the source optimization of SMO. The Level Set Method provides an efficient computational representation of source shape and allows a truly free-form source to evolve during the optimization process.

MEEF is an important factor that must also be considered in SMO for low k1 lithography. When mask and source are optimized without considering MEEF, unrealistic or unusable solutions may result. MEEF and DOF should be considered simultaneously in mask correction and SMO. Use of MEEF in ILT and SMO generates optimized results for low K1 lithography. Edge-based MEEF definition and calculation address the need for more detailed analysis of MEEF in a complex 2D environment in a low k_1 regime. The use of a full-field MEEF color bitmap using edge-based MEEF definition provides easy visual identification of hotspots.

Luminescent's SMO framework provides both Free-form SMO and DOE-constrained SMO, which offers flexible workflow for both memory and logic device patterning. Most logic devices will do better with a DOE constrained source because of random patterns, while memory devices can utilize a free-form source due to their repetitive critical patterns.

Acknowledgments

The authors would like to thank John McCarty and Yoshio Tanaka-san for reviewing and proof-reading this manuscript.

References

1. Abrams, S. and Pang, L., "Fast inverse lithography technology", 31st Internal Symposium of Microlithography, Proc. SPIE 6154 (2006).
2. Hung, C. Y., et al, "Pushing the Lithography Limit: Applying Inverse Lithography Technology (ILT) at 65nm generation", Proc. SPIE 6154, (2006)
3. Lin, B., et al, "Inverse Lithography Technology at Chip Scale", 31st International Symposium of Microlithography, Proc. of SPIE 6154, (2006)
4. Liu, Y., et al, "Inverse Lithography Technology Principles in Practice: Unintuitive Patterns", Proc. SPIE 5992, (2005)
5. Pang, L., et al, "Inverse Lithography Technology (ILT), What is the Impact to Photomask Industry?", Proc. SPIE 6283, (2006)

6. Pang, L., et al, "Inverse Lithography Technology (ILT): Keep the balance between SRAF and MRC at 45and 32 -nm", Proc. SPIE 6730, (2007)
7. Cho, B.U., et al, "Evaluation of Inverse Lithography Technology for 55nm-node memory device", Proc. SPIE 6924, (2008)
8. Pang, L., et al, " Validation of Inverse Lithography Technology (ILT) and Its Adaptive SRAF at Advanced Technology Nodes ", Proc. SPIE 6924, (2008)
9. Prins, S. L., et al, "Inverse Lithography as a DFM Tool: Accelerating Design Rule Development with Model-Based Assist Feature Placement, Fast Optical Proximity Correction and Lithographic Hotspot Detection", Proc. SPIE 6924, (2008)
10. Xiao, G., et al, " Source Optimization and Mask Design to Minimize MEEF in Low K1 Lithography", Proc. SPIE 7028, (2008)
11. Maurer, W., “Mask specifications for 193 nm lithography,” 16th Annual BACUS Symposium on Photomask Technology and Management, Proc. SPIE 2884, 562-571 (1996)
12. Schellenberg, F. M. and Mack, C., “MEEF in theory and practice”, 19th Annual BACUS Symposium on Photomask Technology, Proc. SPIE 3873, 189-202 (1999)
13. Saleh, B.E.A. and S.I. Sayegh, "Reductions of errors of microphotographic reproductions by optical corrections of original masks", Optical Eng., Vol. 20, p. 781-784 (1981)
14. Nashold, K.M. and Saleh, B.E.A., "Image construction through diffraction-limited high-contrast imaging systems: an iterative approach", J. Opt. Soc. Am.A, vol. 2, p. 635 (1985)
15. Hung, C. Y., et al, "Pushing the Lithography Limit: Applying Inverse Lithography Technology (ILT) at 65nm generation", Proc. SPIE 6154, (2006)
16. Lin, B., et al, "Inverse Lithography Technology at Chip Scale, 31st International Symposium of Microlithography", Proc. of SPIE 6154, (2006)
17. Liu, Y., et al, "Inverse Lithography Technology Principles in Practice: Unintuitive Patterns", Proc. SPIE 5992, (2005)
18. Pang, L., et al, "Inverse Lithography Technology (ILT), What is the Impact to Photomask Industry?", Proc. SPIE 6283, (2006)
19. Pang, L., et al, "Inverse Lithography Technology (ILT): Keep the balance between SRAF and MRC at 45and 32 -nm", Proc. SPIE 6730, (2007)
20. Cho, B.U., et al, "Evaluation of Inverse Lithography Technology for 55nm-node memory device", Proc. SPIE 6924, (2008)
21. Xiao, G., et al, "Source Optimization and Mask Design to Minimize MEEF in Low k1 Lithography", Proc. SPIE 7028, (2008)
22. Pang, L., et al, " Considering MEEF in Inverse Lithography Technology (ILT) and Source Mask Optimization (SMO)", Proc. SPIE 7122, (2008)

ECS Transactions, 18 (1) 315-320 (2009)
10.1149/1.3096467 © The Electrochemical Society

Challenges and Opportunities for Optical CD Metrology in Future Lithography Process

Junwei Bao, Jason Ferns, Youxian Wen

Timbre Technologies, Inc.
2953 Bunker Hill Lane, Suite 301, Santa Clara, CA 95054, USA

As the IC industry is progressing towards 32 nm technology, Optical CD metrology will play an important role to help address many challenges in process control. OCD has the unique capability to characterize many critical dimensions in-line for complicated device structures nondestructively. Future process integration requires frequent W2W monitoring and feed forward / feedback control for critical layers. This makes integrated metrology (IM) a key option in process tools, and currently OCD is the only viable option for CD IM. In order to seize these opportunities, many improvements for OCD are need. Optical tools may need to be further optimized for better CD metrology performance. OCD modeling software should be able to model complicated structures and greatly boost ease of use. Furthermore, a well developed IM-SA OCD common platform is needed as the foundation for advanced process and equipment control.

Introduction

Since its debut in semiconductor industry more than ten years ago, optical CD (OCD, a.k.a. scatterometry) metrology has been widely used in many semiconductor fabs for process monitoring and control. Compared to other CD metrology tools, such as CD-SEM, XSEM, AFM etc., it has many advantages: relatively low cost, high throughput, small footprint, which makes it the only technology suitable for CD integrated metrology (IM). In addition, OCD has demonstrated very good precision and capability to measure material interfaces of different refractive indices in complicated structures. All of these advantages make OCD a great candidate for equipment and process control in manufacturing environment.

Over the past decade, OCD has moved from academia to manufacturing. It is used for feed forward (FF) and feedback (FB) control of resist trim and gate etch processes, process monitoring for some very challenging memory structures, litho process control using IM, and so on. However, considering its deployment base in fabs, it is still not a mainstream metrology in semiconductor manufacturing. As the industry is progressing towards 32 nm technology node, many challenges emerge in process development and later in manufacturing ramp up. OCD will be able to play an important role to help address some of these challenges.

In order to seize these opportunities, many improvements for OCD are needed. Both OCD hardware and software need further development to improve system sensitivity, modeling capability, and ease of use. Furthermore, a common IM and SA (stand-alone)

Ex. 1016, Page 343

platform and data management infrastructure need to be developed to fully utilize OCD in manufacturing environment.

OCD Technology Overview

OCD was first commercialized by Sandia System in mid 1990s (1). Custom-designed variable-angle reflectometer was the basic optical configuration. U.C. Berkeley and Timbre Technologies, Inc. demonstrated full profile measurement capability in 1999 using spectroscopic ellipsometer (SE), a mature thin film tool, as optical hardware for OCD (2). Since then, OCD products of various configurations have been released by many metrology vendors for both SA and IM markets, and they have been widely adopted by IC manufacturers (3) (4) (5) (6). For example, Timbre has shipped more than 200 OCD systems to semiconductor fabs worldwide for process monitoring and control. Although OCD needs continuous improvement to meet the needs for future technology nodes, it is widely recognized that OCD will be one of the key metrologies for overcoming difficult challenges in 32 nm nodes and beyond (7).

Similar to thin film metrology, OCD measures fine structures by matching measured optical signal with simulation from a prebuilt structure model (1-2). Today most OCD systems use broadband light sources. SE is the major configuration for SA tools. It has very good sensitivity for most structures, and the technology has been fairly mature due to previous investment on thin film metrology. On the other hand, most IM systems use polarized or depolarized reflectometry design. This is mainly due to the special IM requirements: small footprint, low cost, good reliability and tool-to-tool matching, and fast throughput. One drawback for broadband systems is the need for broadband nk data. Very often nk extraction is a major roadblock for fast recipe setup and obtaining accurate CD results.

Great progress has been made in OCD software in the past ten years. Most OCD systems in production environment use library-based technique due to its fast MAM time and robust results, while regression is often used for model analysis and development. Advanced OCD software is capable to model and simulate optical signal for almost all IC structures; however, in many cases the final measurement results are limited by optical hardware sensitivity. Compared to CD-SEM, it is still more challenging to qualify an OCD recipe. Great improvement is needed in ease of recipe setup to make OCD a true production-worthy metrology.

Current OCD Applications

With the rapid spread of OCD in modern fabs, many applications have been developed. Using OCD full profile data for wafer-to-wafer (W2W) FF and FB control in resist trim and etch processes might have demonstrated the most obvious ROI (Return on Investment). Run-to-run control has been widely adopted in the IC industry since a decade ago; however it mainly uses lot level data by sampling a few wafers per lot due to the equipment cost and measurement delay of CD-SEM. OCD, especially IM, has significantly lowered metrology cost, thus enable W2W control using data from every wafer. In 2006 IBM first published volume production data for APC using CD IM in 90

Ex. 1016, Page 344

nm resist trim and gate etch process (8). Compared to lot-to-lot control by CD-SEM, W2W control by IM has 44% tighter CD distribution, and the percentage of wafers within 1 nm of target is improved from 56% to 88%.

In addition to lateral CD, OCD also provides sidewall angle (SWA), thickness, and other profile information, which makes it possible to do APC with multiple inputs. The same group at IBM showed in 2007 that MISO (Multiple Input Single Output) control models eliminated CD / FB disturbances caused by variation in DI SWA (9). Besides etch time, resist CD and SWA are also used as model inputs for etch bias. Using the MISO model, there was 17% improvement at gate CD 3σ over SISO (Single Input Single Output) model.

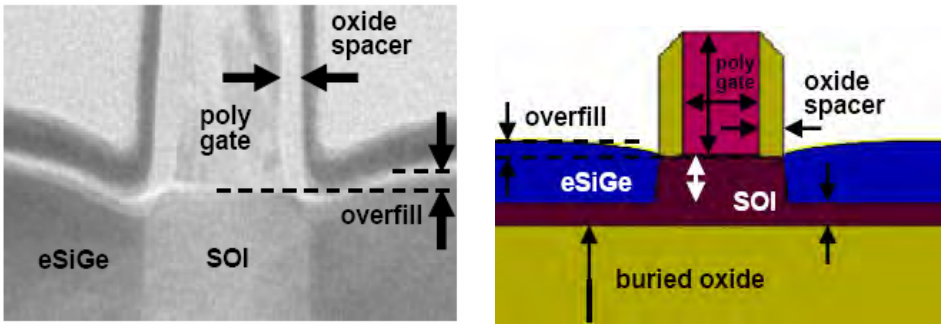


Figure 1. TEM image (left) and OCD model (right) for PFET spacer structure (4)

CD-SEM has been the major workhorse for inline CD metrology in IC manufacturing. However, CD-SEM is usually used to measure interfaces between solid materials and air that are visible from top-down view. For OCD, it is possible to "resolve" interfaces between adjacent solid materials, as long as their refractive indices differ. Figure 1 shows OCD's capability to measure advanced spacer structures with embedded SiGe (4). The ability to measure such structures in-line is critical for prediction of device characteristics which can lead to improved device performance and yield. Experimental data demonstrated good sensitivity for key control features, especially eSiGe proximity and un-etched SOI thickness, which have very tight process control requirements.

For OCD tools integrated on process equipments, there is very little overhead and delay to obtain CD data for all wafers going through the equipment. When using CD-SEM as metrology for SPC, one can only afford to sample a few wafers per lot due to metrology cost, whereas single wafer excursions are normally not detected immediately. With IM, every wafer is measured right after it is processed, therefore allows quick detection of excursion as soon as it happens (3). The massive data generated by IM also contains a lot of information about the process and equipment. If properly analyzed, one can monitor the process and equipment, even in the module level, very effectively without running test wafers.

Future Opportunities for OCD Applications

In order to keep up with the scaling trends, it is essential to have tighter and tighter CD control for critical layers. Naturally the equipment and control strategies that make a process inside the control limits for 65 nm node may not work for 32 nm. One reason for process drift is due to equipment's own instability. As scaling is pushing semiconductor

Ex. 1016, Page 345

process closer to manufacturing tools' physical limits, some minute drifts in tool conditions might be magnified for devices printed on wafer and make it fall out of process window. Even if equipment vendors can make future-generation tools more stable, the stability of a process can also be limited by the stability of incoming stack. For example, SiON BARC thickness variation can cause similar amount of CD variation for 45 nm litho process. In future technology nodes, the impact of CD variation due to BARC thickness variation could become significant due to shrinking CD budget. In addition to relying on improvements from equipment vendors, IC manufacturers can implement their own APC solution using OCD data to compensate variations of litho tools and incoming layers.

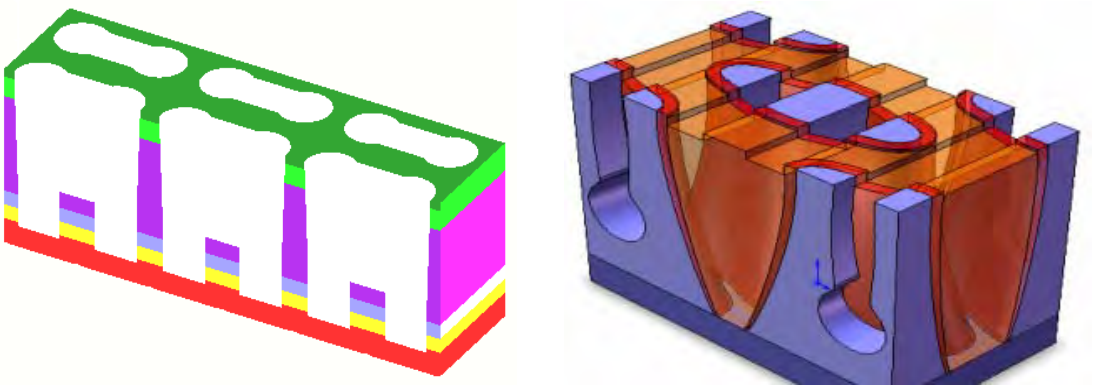


Figure 2. Example memory structures for OCD applications

As device scaling is approaching material physical limits, innovative device structures are used for performance and density enhancement to keep up with Moore's Law. When developing production-worthy solutions for these new devices, very often it is necessary to monitor and control process quality of some complicated 3D structures. Figure 2 illustrates two memory structures that need to be measured by OCD. Some critical dimensions are between two solid materials and / or buried under other layers. Cross-section SEM was the only metrology for these structures before, and now OCD becomes the key enabler for such processes in volume manufacturing.

As demonstrated in references (9) and (10), resist SWA is becoming more important in etch CD bias control. Furthermore, other detailed profile information, such as footing, sidewall bowing, etc., may be important for patterning process control and device performance. As the only viable full-profile metrology for APC in volume production, OCD will play a key role as fabs' control framework evolves from SISO to MISO to MIMO (Multiple Input Multiple Output).

Pitch splitting has become an essential process to keep up with the scaling trends before EUV is ready for volume production. A wide range of pitch splitting techniques have been proposed, many of which have CD-overlay confounding issues and need tight control of both CD and overlay. OCD has emerged as a very promising metrology being able to precisely measure CD and overlay simultaneously.

Due to its excellent performance in throughput and precision, OCD has also been used in process equipment qualification and tuning (11). Bake plate optimization for a litho track using OCD showed that WiP (within plate) uniformity improved by 57%, and PtP (plate to plate) uniformity by 93% for L&S patterning process on base Si. Over 50% improvement for PtP uniformity was demonstrated on production wafers also. There was

Ex. 1016, Page 346

work reported on scanner qualification using OCD as well. As more work on process tool fine tuning and quick qualification is needed in the future, OCD will be more widely adopted in these applications.

Gaps for OCD Technology to Meet Future Needs

As an emerging technology, many improvements are needed for OCD to catch these opportunities. An essential task is to continue improving its performance on smaller CDs, fine features in complicated structures, and overlay. Although advanced OCD software can model most device structures, the critical dimensions may not be detectable by the optical signals. In order to get better sensitivity to these features, OCD hardware needs to have better signal-to-noise ratio, extend to broader wavelength range, and possibly combine multiple optics for more "signal content." Most popular OCD systems today evolved from thin film metrology tools due to similar operation principles and maturity of these tools. However, as OCD requirements tighten, it becomes essential to fully optimize existing hardware for OCD, or even completely redesign the optics, to achieve better system accuracy and tool-to-tool matching in the future. Some of these next-generation hardware platforms are already under beta-test at customer sites.

The most frequent complaint to OCD product is the difficulty to set up a recipe. Compared to thin film metrology, OCD significantly extends the modeling space, and the profile parameters often have relatively weak sensitivities and high correlation. Furthermore, computation takes orders-of-magnitude longer time than that for film stacks. Therefore it frequently takes more effort to build an OCD model than film applications, though significant progress has been made in the past 5 years. For OCD to become a mainstream metrology in the production environment, more progress is needed for automated model generation, so that a process engineer or operator can generate OCD recipe instead of counting on metrology engineers with a lot of OCD experience. In addition, advances in handling nk variation and computation speed up for complicated 3D structures are also key to the wide adoption of OCD technology.

One of the great potential benefits for OCD is its capability to share a common platform between IM and SA. Once this is realized, IM and SA data can be used interchangeably. One can use SA systems for initial recipe and process development, and then transfer the recipe to IM. During process ramp up phase, very high sampling is needed to "learn" and closely monitor the process. At this time SA tools can be used to measure certain portion of wafers, complementing IM's bandwidth. Once the process gets stable and enters volume production phase, only IM is needed for process monitoring and control. While SA tools with additional optics may be still used occasionally for diagnostic purpose, their major capacity can be moved over to other process development.

There has been a wide spectrum of OCD software products and optical hardware configurations developed: SE, broadband normal incidence or laser-based variable-angle reflectometers with polarization or depolarization. Even for the same configuration, e.g. SE, the design for different hardware models could be quite different. While it is good for an emerging technology to try different options, there has not been enough effort to design a common product platform for both IM and SA. Fortunately some equipment and metrology vendors started joint development efforts for common platform recently, and we should expect such products being released in the near future.

Ex. 1016, Page 347

Summary

Many valuable OCD applications have been developed in the past few years. The difficult challenges in future technology nodes provide great opportunities for OCD to extend its application space. Improvements needed for OCD technology are analyzed, and new technologies are being developed by OCD vendors to close these gaps.

References

1. C. Raymond, "Overview of Scatterometry Applications in High Volume Silicon Manufacturing," *2005 International Conference on Frontiers of Characterization and Metrology for ULSI Technology* (2005).
2. X. Niu, N. Jakatdar, J. Bao, C.J. Spanos, S. Yedur, "Specular Spectroscopic Scatterometry in DUV Lithography," *Metrology, Inspection and Process Control for Microlithography XIII*, Proc. **SPIE 3677**, 159-168 (1999).
3. U. Kramer, G. Fleischer, T. Marschner, S. Hornig, H. Weichert, D. Hetzer, "Application of integrated scatterometry measurements for a wafer-level litho feedback loop in a high-volume 300 mm DRAM production environment," *Data Analysis and Modeling for Process Control III*, Proc. **SPIE 6155**, [6155-27] (2006).
4. M. Sendelbach, *et al*, "Assessing scatterometry for measuring advanced spacer structures with embedded SiGe," *Metrology, Inspection and Process Control for Microlithography XXII*, Proc. **SPIE 6922**, [6522-25] (2008).
5. O. Sorkhabi, *et al*, "Performance of measuring contact holes using the Opti-Probe 7341 3-D RT/CD technology," *Metrology, Inspection and Process Control for Microlithography XIX*, Proc. **SPIE 5752**, 912-921 (2005).
6. R. Sadjadi, *et al*, "Plasma-assisted CD shrink and overlay metrology techniques for double patterning," *2007 International Symposium on Semiconductor Manufacturing*, 1-3 (2007).
7. Semiconductor Industry Association, "Metrology", *International Technology Roadmap for Semiconductors, 2007 edition* (2007).
8. M. Sendelbach, *et al*, "Integrated scatterometry in high-volume manufacturing for polysilicon gate etch control," *Metrology, Inspection and Process Control for Microlithography XXI*, Proc. **SPIE 6152**, [6152-15] (2006).
9. K. Bandy, E. Meyette, B. Hampton, D. Prager, H. Lee, M. Funk, A. Yamashita, "Polysilicon Gate Etch Process Control using Integrated Scatterometry and Multiple-Input Wafer-to-Wafer Process Control," *AEC/APC Symposium* (2007).
10. S. Ali, *et al*, "Application of integrated scatterometry (iODP) to detect and quantify resist profile changes due to resist batch changes in a production environment," *Metrology, Inspection and Process Control for Microlithography XXI*, Proc. **SPIE 6518**, [6518-94] (2007).
11. K. Ruck, *et al*, "Effect and procedures of post exposure bake temperature optimization on the CD uniformity in a mass production environment," *Metrology, Inspection and Process Control for Microlithography XXI*, Proc. **SPIE 6518**, [6518-192] (2007).

ECS Transactions, 18 (1) 321-332 (2009)
10.1149/1.3096468 © The Electrochemical Society

Line Edge Roughness Analysis for KrF Photoresists under Sub-0.11 um Groundrules

Lei Wang, Liang Ji, Peng Wu, Huan Kan, Weimin He, Zhiguo Zhu, Fucheng Chen, Fang Wei, Jun Zhu

Shanghai Huahong NEC Electronics Company, Ltd.
No. 1188 Chuanqiao Road, Shanghai 201206 China

To maintain cost competitiveness for the 200 mm wafer production, the industry starts to explore production feasibility under groundrules smaller than 110 nm while maintaining the cost advantages given by the KrF exposure tools systems. The k1 factor under 110 nm will be below 0.35. It has the similar level of complexity in optical proximity correction compared to 65 nm at 0.93 NA with 193 nm exposure tools. However, due to that the KrF photoresists are originally designed for ground rules greater than 110 nm it is relatively unknown how feasible it is to extend their performance below 110 nm. As we know, in each photoresist design, there is a balance among three major performance indices, which are spatial resolution, photospeed, and pattern edge roughness. For example, a fast photoresist may not have the same resolution or pattern edge roughness performance as a slow photoresist, which may, of course, has low productivity. A good production resist usually has a good balance of all three parameters. This paper will explore the line edge roughness parameters under sub 110 nm design rules and will explore the performance dependence on imaging conditions. Classical photo process optimization result is shown and acceptable LWR performance is achieved. In addition, a comparison in line edge roughness among 3 production KrF photoresists will be made.

Introduction

The continuing groundrule shrink in semiconductor manufacturing has made the linewidth control increasingly challenging. Compared to the relatively known knowledge associated with linewidth shrink, the knowledge associated with line edge roughness (LER) or linewidth roughness (LWR) is relatively unknown. In order to improve the LER performance, people had made some extensively study in past years. In general, LER is related to factors such as, image contrast, resist formulation, process condition and so on. Although the use of shorter wavelength is an effective way to print smaller features, the process cost will become higher with more expensive exposure tools and materials. Therefore for foundry companies, quite a few start the exploration on the extending of KrF process capability to below sub-0.11um groundrule. (1) LER, together with other key process parameters, such as process windows and mask error enhancement factor (MEEF) become a focus of such studies.

Ex. 1016, Page 349

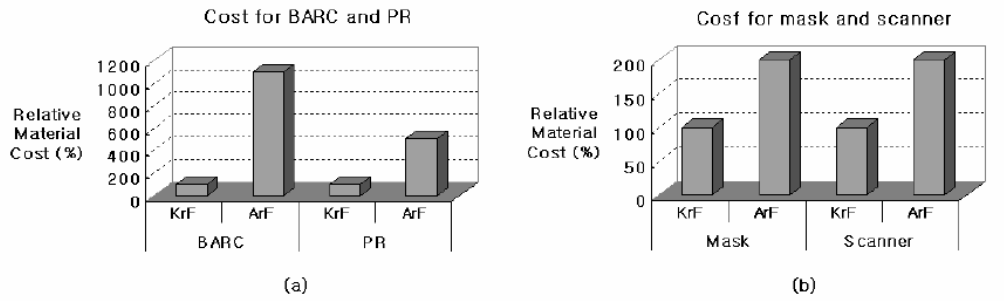


Figure 1. If KrF lithography process manufactures 90nm technology devices, the cost of fabrication can reduce. (a) cost for BARC and PR (b) cost for mask and scanner.

But in general, it is a common belief that KrF process can not meet the sub-0.11um technology node requirement due to the k1 is too low and very difficult for process control especially for LWR control (2). In addition, an important reason for such belief is that the original KrF photoresists are designed for ground rules greater than 110 nm it is relatively unknown how feasible it is to extend their performance below 110 nm. As we know, in each photoresist, there is a balance among three major performance indices, which are spatial resolution, photospeed, and pattern edge roughness. A good production resist usually has a good balance of all three parameters. Therefore, the key to a good process under 0.11 um groundrule is to find a good photoresist and to find an optimized process.

In this paper, we will show our study on LER performance and LER improvement at gate poly layer of 90nm technology node. First, the classical photo process optimization result is shown and acceptable LWR performance is achieved. Second, a comparison in line edge roughness among 3 production KrF photoresists is made. Our studies will also demonstrate some typical performance in different behaviors under bright field and dark field (3).

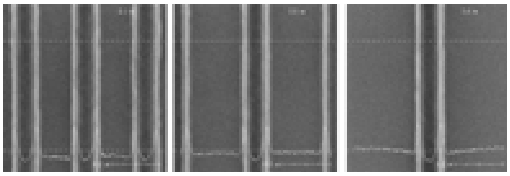
Experiment and Metrology Method

We studied poly layer lithography at 90nm technology node. The target feature design follows generic 90nm logic process. A KrF scanner is used as an exposure tool with maximum NA 0.82 and maximum sigma-out 0.88 and maximum annular ratio 0.75. For a minimum pitch of 0.25 um and a maximum NA of 0.82, masks are fabricated using attenuated PSM (Phase Shift Mask). Three KrF chemically amplified positive tone resists are used. We have performed optimization in the illumination method, PEB (Post Exposure Bake) temperature and mask bias.

.

The choice of CD metrology method is also another important factor to LER evaluation. For 90 nm gate technology node, Hitachi 9300 series CD SEM is used with 200K magnification. In metrology, we will measure LWR (Line width roughness), LER-L (Line edge roughness – left) and LER-R (Line edge roughness - right) for each condition. Three structures are audited (dense, semi-dense at forbidden pitch, isolated line), different shots, different locations in one shots, horizontal and vertical line are also monitored in the study. Typical top down images are shown in Fig. 2:

Ex. 1016, Page 350



(a) (b) (c)

Figure 2: (a) Dense features (b) Semi-dense feature (c) isolated feature

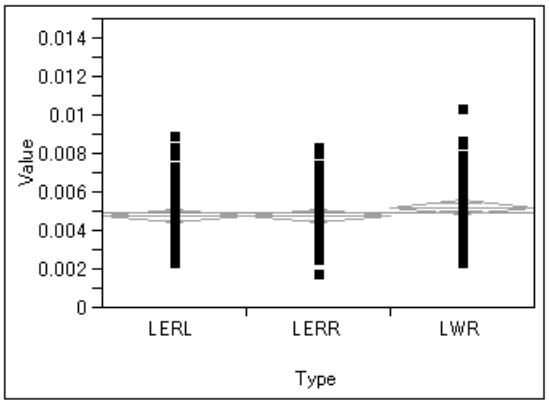


Figure 3: LWR/LER-L/LER-R correlation

Fig. 3 shows the correlation between LWR/LER-L/LER-R. From this figure, we can find that the LWR has higher sensitivity compared to just LER-L or LER-R. The mean value and 3-sigma of LWR are larger than LER-L and LER-R, which is mathematically understandable. Based on the statistical result we will focus on LWR in the next step to for optimization. The LWR average value and LWR uniformity are all important for us.

To achieve our target, we want to control the all structures LWR mean and 3 sigma value to within 5 nm, which is consistent with ITRS requirement for 90nm technology node.

Classical Photo Process Optimization Result and Discussion

1. Illumination Conditions

As we all know, the imaging resolution is critical to LWR. If the resolution is not enough the line profile will become worse and it is very difficult for LWR control. Changing the illumination conditions is a straight forward method to improve LWR through improving image contrast. Because the balance for resolution, photospeed and LWR need based on enough resolution, so we will check MEEF (Mask Error Enhancement Factor) at the same time.

Because KrF process image contrast in dense features is lower than that of ArF process due to exposure wavelength difference, higher NA and OAI (Off Axis Illumination) are required to achieve better image contrast. But too high NA and too strong OAI will reduce process window for the isolated features. In order to balance image contrast for dense features and process window for isolated features, we have tried

Ex. 1016, Page 351

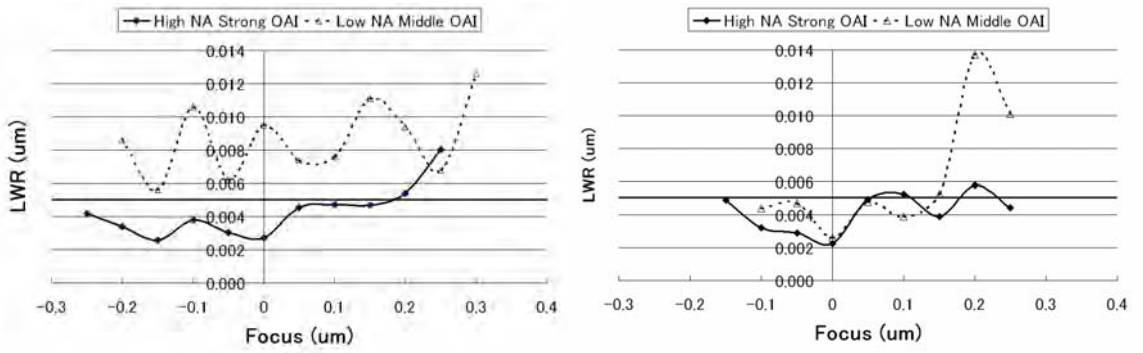
two options for this layer. One is high NA and strong OAI and the other is low NA and medium OAI.

At the same time, as we all know, we will determine an optimized condition for all splits to avoid process variation impact.

1.1 Process Window and CD Proximity. Displayed in Table.1 are summary of process windows for dense and isolated features under high NA strong OAI and low NA medium OAI conditions. The result shows that high NA and strong OAI can achieve better performance of dense features and reasonable performance of isolated features. For non-scattering bar process, 0.22um DOF @ 7% EL for isolated feature is acceptable.

	Dense	ISO
High NA Strong OAI	0.41 @ 8% EL	0.22 @ 7% EL
Low NA Middle OAI	0.33 @ 8% EL	0.22 @ 7% EL

Table 1: DOF (in um) comparison between high NA strong OAI and low NA



(a) (b)

Figure 4: (a) LWR comparison for dense features between high NA strong OAI and low NA medium OAI depending on focus variation at poly layer (b) LWR comparison for isolated features between high NA strong OAI and low NA medium OAI depending on focus variation at poly layer

The LWR dependence on focus variation is compared for two processes shown in Figs.4(a) and 4(b). In LWR, the high NA strong OAI process is better in through focus LWR control.

We have also checked the through pitch LWR at the same time as we study in process window analysis. Here for features at all pitches from dense to isolated, we want to control all features LWR to within 6 nm (5 nm is our improvement direction). The result is shown in Fig. 5. In Fig.5 we can see almost all features LWR can be controlled well for high NA strong OAI and critical features can be controlled to within 5 nm. It is consistent to our result in process window analysis.

Ex. 1016, Page 352

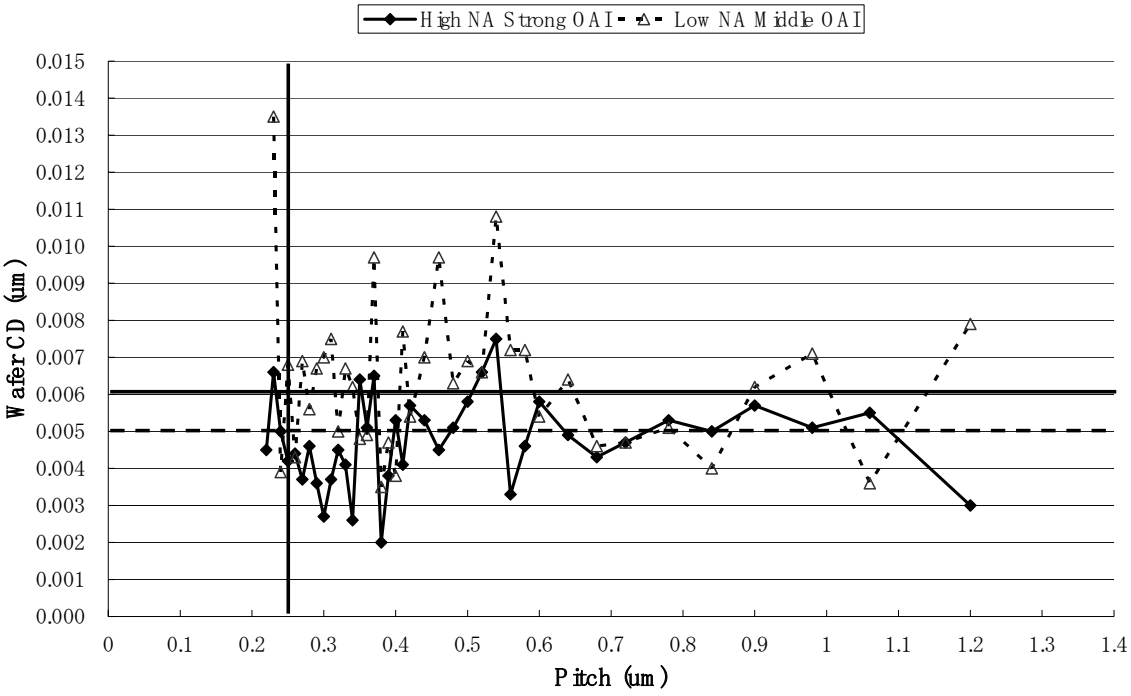


Figure 5: LWR through pitch comparison between high NA strong OAI and low NA medium OAI

Applying the result from the illumination optimization, we will use high NA strong OAI for the experiments in next step to achieve better LWR.

1.2 MEEF Result and Discussion. But which factor is most sensitive to LWR performance difference for these two conditions? We will check the MEEF for dense features. Fig. 6 shows that the experimental ADI CD results for main features with line width from 0.10um to 0.12um. In the figure, the MEEF under high NA strong OAI is 2.08, which is much better than that under low NA medium OAI, around 2.78. It is consistent with our normal understanding that high NA and strong OAI will improve MEEF since high NA strong OAI condition has better image contrast.

The illumination split experiment shows strong relationship between MEEF and LWR. As smaller MEEF is found associated with high NA and strong OAI, we can control the CD uniformity and LWR more easily from photo process side. In future process improvement we will fine tune the other conditions to improve MEEF. On the other side, we will also check the correlation between MEEF and LWR to find better balance.

Ex. 1016, Page 353

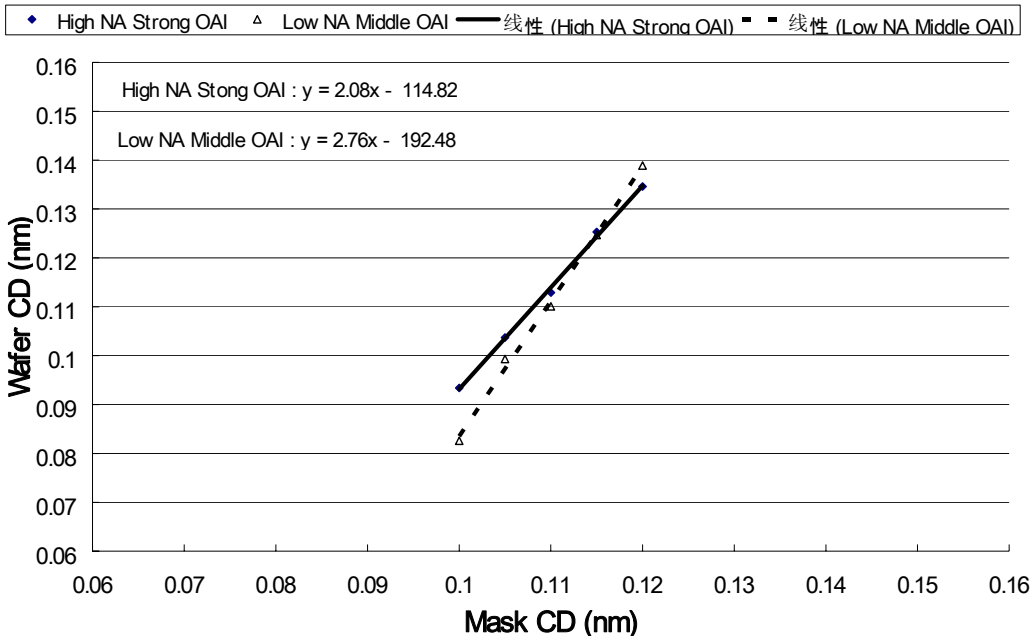
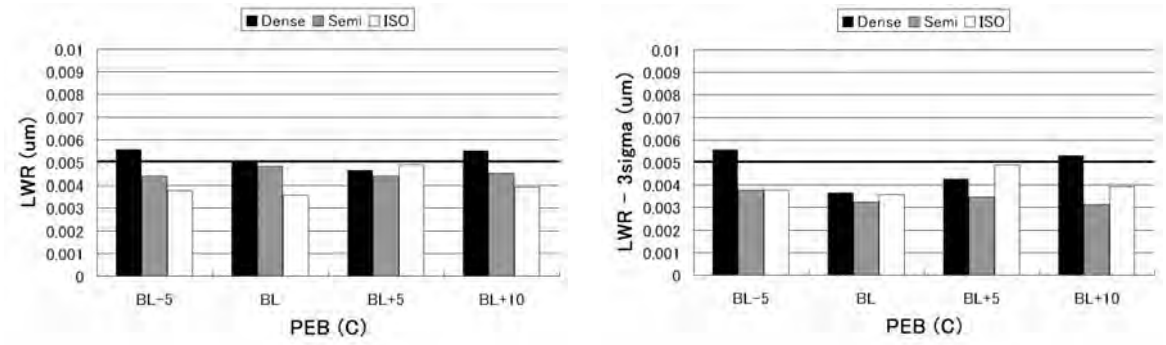


Figure 6: MEEF comparison between high NA strong OAI and low NA medium OAI

2. Post Exposure Bake Conditions

As we all know, bake conditions is also very important to KrF photoresists performance. Normally due to photoacid diffusion, post exposure bake (PEB) temperature variation can affect the resist profile, process window and photospeed by impacting diffusion length. Normally lower PEB will have better resolution but worse profile and higher PEB will have better profile but worse resolution. So the key is to find optimized PEB temperature.

We have tried a total of 5 PEB temperatures ranging from -5 to +15 degrees added on top of the baseline setting. The LWR mean and 3-sigma result is shown in Fig. 7.



(a) (b)

Figure 7: (a) LWR mean value comparison of different PEB temperatures (b) LWR 3sigma comparison of different PEB temperatures

From Figs. 7(a) and 7(b), we can find that there is relatively little trend for LWR versus PEB change. The baseline PEB temperature is actually better than the other conditions. It is a very interesting phenomenon.

To further study this effect, we have checked MEEF for each condition again. The result is shown in Fig. 8.

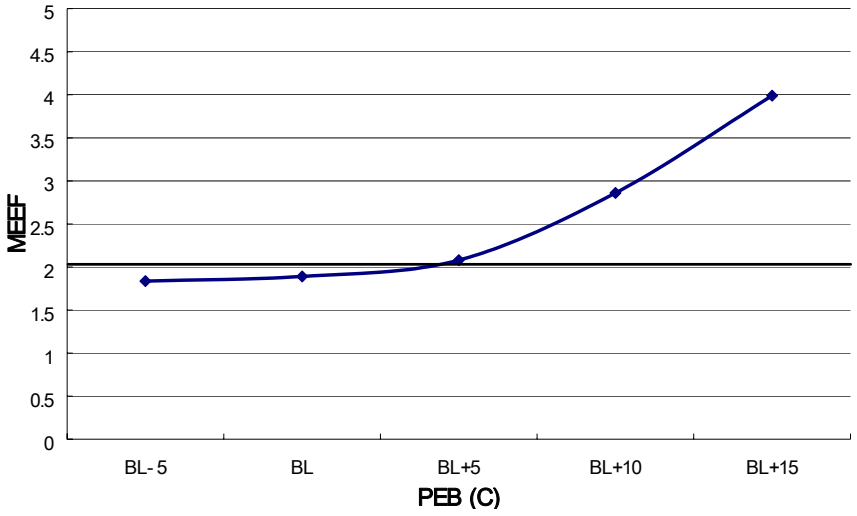


Figure 8: MEEF Vs PEB temperature

	Dense	Semi	ISO	Profiel	MEEF	LWR - Mean	LWR - 3sigma
BL -5C				☹	☺	😐	☹
BL				😐	☺	☺	☺
BL +5C				☺	😐	☺	😐
BL +10C				☺	☹	☹	☹

Table 2: PEB temperature split summary

From the MEEF check result shown in Fig. 8, we can find that the LWR behaves differently from MEEF. With lower MEEF, baseline PEB conditions achieve better LWR mean value and uniformity although the profile is not as good as higher PEB temperature. But if PEB temperature decreased to a specific temperature, the MEEF is comparable for all PEB conditions. In this case, higher PEB would be better in LWR control for better profile.

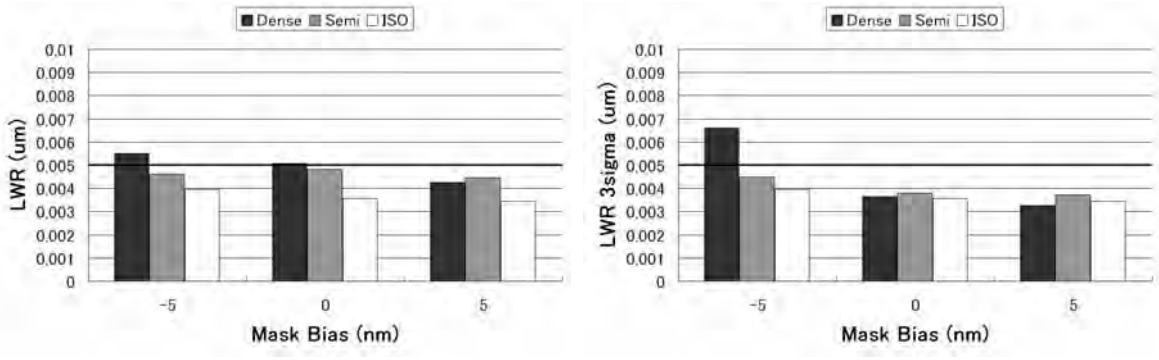
Ex. 1016, Page 355

The result and top view images we have achieved are summarized in Table 2. This summary indicates that the MEEF is a key factor that influences the final LWR control. Smaller MEEF can get better LWR mean value and uniformity even with relative worse profile. On the other side, profile control also could be reflected in LWR performance. If the LWR is not good enough, the LWR will become worse especially in uniformity control.

3. Mask Bias

When the illumination setting and PEB temperature is fixed, mask bias will be another important factor to balance LWR and photo process performance.

We try different mask bias in our study. Since we notice that neither too much over-exposure nor too much under-exposure is good for process windows, we limit ourselves to the use of only +/-5nm mask bias for our study. The result is shown in Fig. 9.



(a) (b)

Figure 9: (a) LWR mean value comparison of different mask bias (b) LWR 3sigma comparison of different mask bias

From the result we can see that mask bias can impact the LWR though the impact does not show the same magnitude as some other factors. Although the difference in LWR mean value is not so obvious, LWR uniformity can still be improved with mask bias optimization. At the same time, mask bias can also improve the MEEF performance. The MEEF check result is shown in Fig. 10. We can see that MEEF is consistent to LWR result. It also predicts that control MEEF is effective to LWR improvement.

The total result and top view image is summarized in Table 3. From this summary, we can make the same conclusion just as we get in the NA/PEB temperature split. MEEF is a very key factor related to final LWR performance. It shows a good method for us to reduce LWR when we setup the process.

After we optimize our process, we can control all structure LWR mean value to within 4.2 nm and the 3 sigma to within 3.7nm. These results are acceptable for 90nm gate poly layer LWR requirement according to general device and product.

Ex. 1016, Page 356

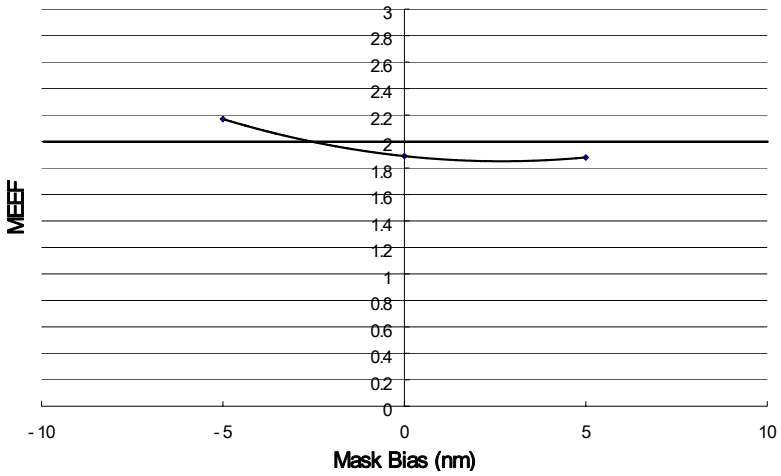


Figure 10: MEEF vs mask bias

	Dense	Semi	ISO	Profiel	MEEF	LWR - Mean	LWR - 3sigma
BL -5nm				☹	😐	☹	☹
BL				😐	☺	😐	☺
BL +5nm				☺	☺	☺	☺

Table 3: Mask bias split summary

Results and Discussion for Different Photoresists

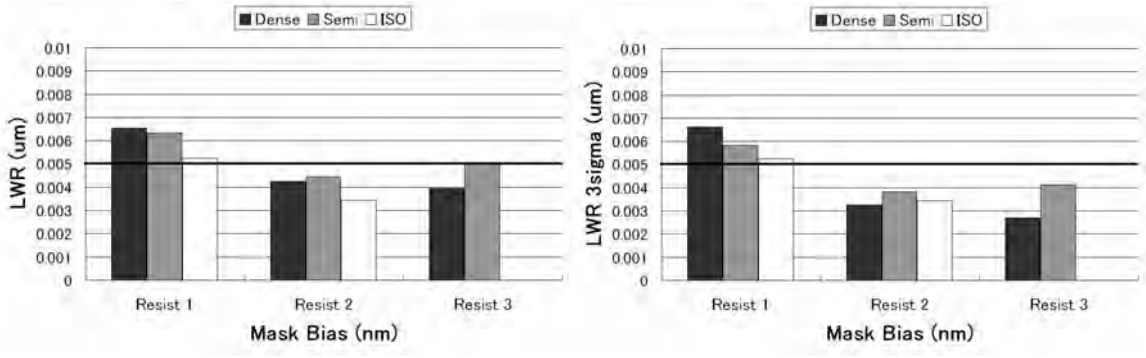
All the optimization and check items are standard in photolithography. With the selection of a good photoresist, we have demonstrated that we can optimize our process to realize balanced performance for process window as well as LWR. Therefore the key to the success of a sub-0.11 um photo process with KrF exposure tool is to find a suitable photoresist.

According to current popular diffusion model of photo resist, the process window, LWR, MEEF, photo speed and all kinds of photo resist characteristics are related to photo acid diffusion of the aerial image during post exposure bake (PEB). Resist formulation designers have lots of methods to change the resist performance. Then from the final performance, what they do can be summarized as how to control the photo acid efficiency, diffusion behavior, chemical contrast, defectivity performance, temperature sensitivity, stability and so on. According to the general classification, all photoresists can be categorized into low activation energy (or low temperature), high activation energy (or high temperature), or hybrid (a combination of both low activation and high

Ex. 1016, Page 357

activation components). Since a good photoresist must be capable of exhibiting good LWR, process window, and low exposure dose and so on, we can characterize the resist performance by simple process behavior [3]. Here we will focus on the difference in LWR.

In order to make the phenomenon more clearly, we will use the same resist in our previous study for bright field and dark field study [3]. The first resist, or resist 1, is a typical line resist used in 0.13um or above technology node gate layer. Normally this kind of resist will show good MEEF control, balanced CD proximity, and better pattern collapse control. But the resolution and LWR is not so good for tight pitch. The second resist, or resist 2, is a typical metal trench resist used in copper BEOL process. It will show good resolution especially in space pattern to avoid metal bridging and good profile control. But it will show worse CD proximity control such as resist 1. Finally, resist 3 is a typical contact hole resist. Although it will show very good dissolution contrast, too high dissolution contrast to aerial image is actually not good for semi- to isolated line patterns, mostly under bright field condition. It will show bad performance in isolated line-pattern and more susceptible to pattern collapse.



(a) (b)

Figure 11: (a) LWR mean value comparison of different resists (b) LWR 3sigma comparison of different resists

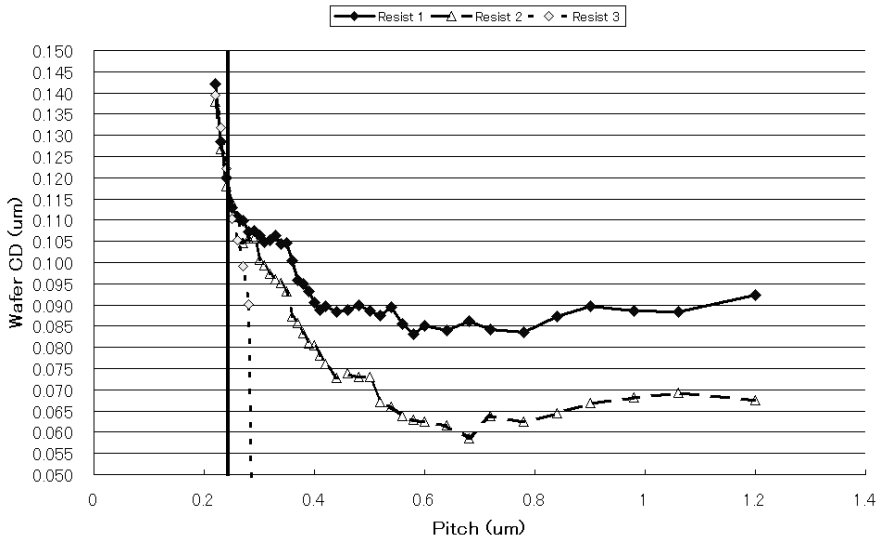


Figure 12: CD proximity of different resists

Ex. 1016, Page 358

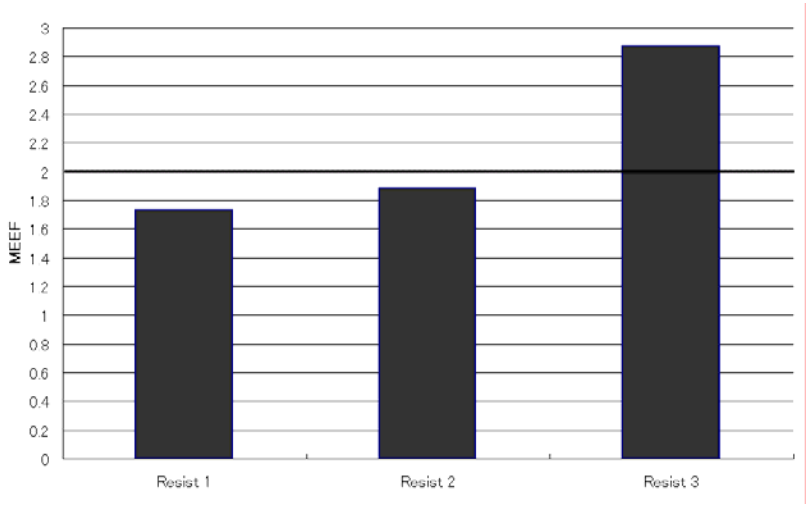


Figure 13: MEEF of different resists

The LWR experiment result is shown in Fig. 11, CD proximity is shown in Fig.12, and MEEF result is shown in Fig. 13. The result is consistent with our general understanding. At tight pitch, although the MEEF performance of resist 1 is good but it still shows bad performance in LWR especially in uniformity control. Because its profile is not good as we want. Although the MEEF and process window is good, it still does not show good performance in LWR control. For resist 3, although the MEEF is high, it is very good at target feature. But it has no process window for isolated features even it has big window for dense features since it is too sensitive to all kinds of condition variation under bright field. However, its proximity is unsatisfactory and not acceptable for poly layer. Finally resist 2 shows acceptable process control capability at the same time when it exhibits good LWR control capability. Besides above result, another interesting observation is that the MEEF result is also consistent to our result in the last chapter. First, we need to control MEEF to an acceptable level before we optimize our process for LWR control. It also can help photo process engineers when they challenge the process limitation at low k1 conditions. The summary table is shown in Table 4.

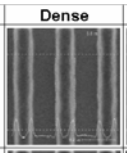
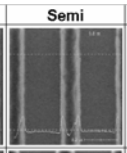
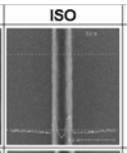
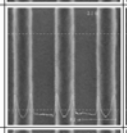
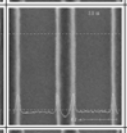
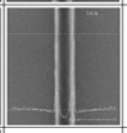
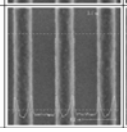
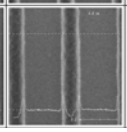
	Dense	Semi	ISO	Profiel	MEEF	LWR - Mean	LWR - 3sigma	Common DOF	ID Bias
Resist 1				☹	😐	☹	☹	☺	☺
Resist 2				☺	☺	☺	☺	😐	😐
Resist 3			110nm can not be printed without scattering bar	☺	☹	😐	😐	☹	☹

Table 4: Resist split summary

Finally we conclude that resist 2, a typical space resist, is the best among the 3 resists we have studied. At the same time, we believe that this study can also serve as an important reference for not only KrF dry but also ArF dry and immersion processes.

Ex. 1016, Page 359

Conclusions

We have performed a LER study of KrF process at gate poly of 90nm technology node. Classical photo process optimization method is used. We get mean within 4.2nm and 3-sigma within 3.7nm for dense, semi-dense and isolated features. It is an acceptable result for a typical 90nm logic process. Besides, we have also discussed the LWR different from 3 different photoresists.

We expected KrF process can get enough process control capability for this generation to instead of ArF process to reduce the cost although current KrF scanners with maximum 0.82 NA. Although the aerial image contrast from KrF under 90 nm groundrule can not compare with that from ArF scanners, due to the fact that KrF photoresists are more mature, the combination, as we have demonstrated, can produce enough process window for a 90 nm process. Consequently, we believe that we can apply KrF in sub-0.11um technology node by new higher NA KrF scanners, optimizing the process and improving the process margin.

Acknowledgments

The author of this paper would like to thank the production division, the technology division, the third engineering division, the design service division, and the management teams of Shanghai Huahong NEC Electronics Company, limited for their kind support of the engineering work for this paper.

References

1. Jo Finders, et. al., *Can DUV Take Us Below 100nm*, Proc. SPIE **4346,** pp153 (2001).

2. Shangho Lin, et. al., *How to Print 100nm Contact Hole with Low NA 193nm Lithography*, Proc. SPIE **5376,** pp1091 (2004).

3. Lei Wang, et. al., *The Characterization of Photoresist for Accurate Simulation Beyond Gaussian Diffusion*, Proc. SPIE **6519-35** (2007) and references therein.

ECS Transactions, 18 (1) 333-339 (2009)
10.1149/1.3096469 © The Electrochemical Society

Lithography Process Transfer from ArF to High-NA KrF Tool

Jongkyun Hong[a], Alexander Serebryakov[b]

[a]ASML, No.59 Keji 6th Rd, Kuei-Shan, Taoyuan, Taiwan,
[b] ASML, De Run 1110, 5503 LA Veldhoven, The Netherlands
Contact: jk.hong@asml.com

In this paper, we present the capabilities of lithographic process transfer by use of high NA KrF tool (TWINSCAN™ XT: 1000H, Maximal NA 0.93). The Metal 1 (100 nm half-pitch) and Via (110 nm half-pitch) layers of 65 nm technology have been chosen for this research. The layers are from a product that is currently manufactured on ArF tool (TWINSCAN™ XT: 1400). By optimizing the illumination setting and applying an advanced OPC solution, we have obtained the overlapping process window of critical features for the KrF process that is comparable with parameters of the initial ArF process. Good performance of Metal layer with respect to MEEF (Mask Error Enhancement Factor) also allows for stable process characteristics within wide line through pitch range. However, to ensure CD uniformity for Via layer an aggressive illumination solution such as soft-QUASAR™ or customized illumination may be required.

Introduction

In order to minimize manufacturing costs, lithographers have to extend capabilities of i-line and KrF tools working with a low k1 factor. In 2008, ASML shipped the first TWINSCANTM XT: 1000H – a new KrF tool with the highest NA (0.93) on the KrF market. For ASML customers, this tool offers new opportunities for improving value of ownership.

To demonstrate the high performance capabilities of TWINSCAN XT:1000H, we have researched the feasibility of transferring Metal 1 (100 nm half-pitch) and Via (110 nm half-pitch) layers of 65 nm customer product currently produced on XT:1400 to the XT:1000H. In order to meet low-budget requirements, we use only standard illumination modes (annular, conventional and quasar). Another constraint is not to use polarization options present on XT: 1000H.

By optimizing the illumination settings and applying advanced mask design approaches with help of ASML litho imaging software, we demonstrated the possibility of significant extension of KrF litho process margin with XT:1000H even in presence of substantial limitations noted above. The achieved process margin of KrF process meets all customer specifications with respect to exposure latitude and depth of focus.

However, the MEEF value for critical pitches of Via layer indicates forbidden pitch region that requires special approach using customized illumination mode or design layout relaxation for improving wafer CD uniformity.

Ex. 1016, Page 361

Methodology and Simulation Setup

Methodology

To achieve successful process transfer, the following 3 step process flow has been utilized. First we select characteristic features that indicate the process performance for different contrast threshold e.g. set of CD through pitch structures, and then we calculate the reference process window for them. These characteristic features are used as the reference for optimizing the illumination setting and mask design. At this optimization stage, the size of common process window is used as criterion. Then, qualification is performed by comparing performance of secondary process with respect to the original one in terms of overlapping process window. The MEEF value cross check is discussed below in separate section.

- Step 1. Check reference process capability and select characteristic features to determine weak points within layout of mask design. These features indicate critical contrast values that can cause imaging challenges.
- Step 2. Optimize illumination setting and mask design with process window boundaries that satisfy the specification.
- Step 3. Qualify the process transfer feasibility by comparing process performance on the original and secondary tool with overlapping process window and also MEEF.

Simulation Setup

To show the process transfer feasibility, aerial image based simulation with LithoCruiser™ has been performed. The simulation conditions have been defined as follows:

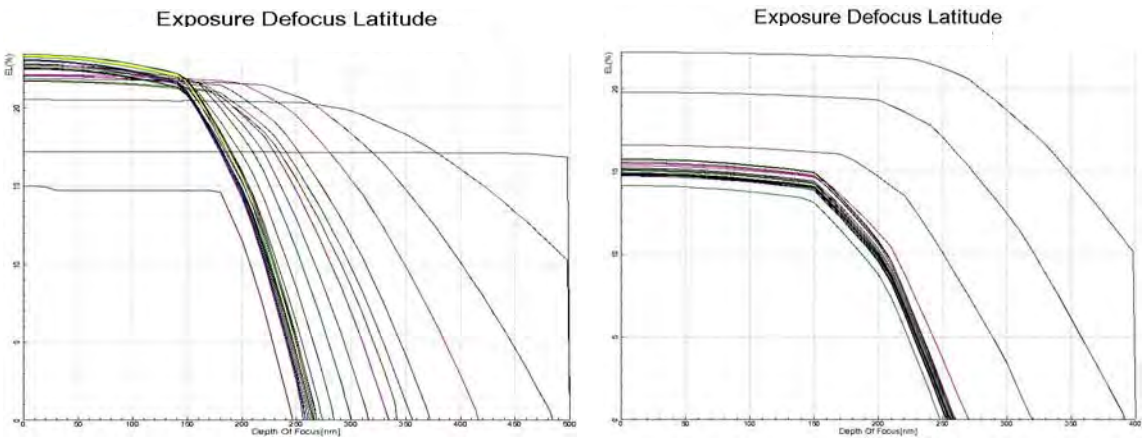
- Simulation engine : LithoCruiser™ with NA/sigma optimizer and rule based OPC modules
- Process transfer criteria : comparable to reference process window
- Process transfer condition : limit as standard illumination mode
- Mask type : 6% attenuated phase shift mask
- Feature size : 100nm half-pitch line and space features through pitch
  110nm half-pitch 2 dimensional space features through pitch

Simulation Results

(1) Reference Process Window and Characteristic Feature Selection

Reference Process Window. Using the current production exposure setting for the ArF system, the reference process windows for CD through pitch reference structures have been simulated (see Figure 1). Results shows that the common process window for Metal layer is of 15% exposure latitude with 250nm depth of focus and for Via layer of 14% exposure latitude with 250nm depth of focus.

Ex. 1016, Page 362



(a) Metal layer (b) Via layer

Figure 1. Reference ArF process window: metal layer (a) and via layer (b).

Characteristic Feature Selection. Considering proximity effect for CD change through pitch, the characteristics features have been selected for both of Metal and Via layer. The Metal layer's proximity curve and critical features are shown in Figure 2. These features represent the proximity behavior with good accuracy. Five features in Metal layer and four features in Via layer have been selected from dense to isolated feature type. These are used for optimizing illumination setting and mask design.

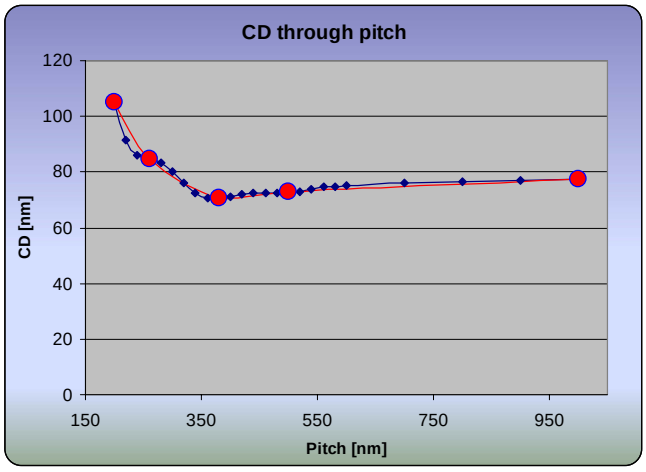
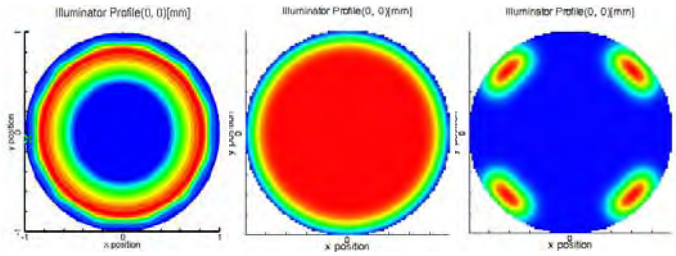


Figure2. CD through pitch curve with characteristic features (large circles in the graph)

(2) Illumination Setting and Mask Design Optimizing

Illumination Setting Optimization. As already noted, the optimization of illumination mode is only allowed for standard illumination schemes. Using automatic NA/sigma DOE optimization module in LithoCruiser^{TM}, the illumination setting have been optimized for maximizing overlapping process window with original mask design as shown in Figure 3. As a result, annular illumination mode has been recommended for Metal layer and in case of Via layer, conventional mode and also quasar illumination mode have been proposed.

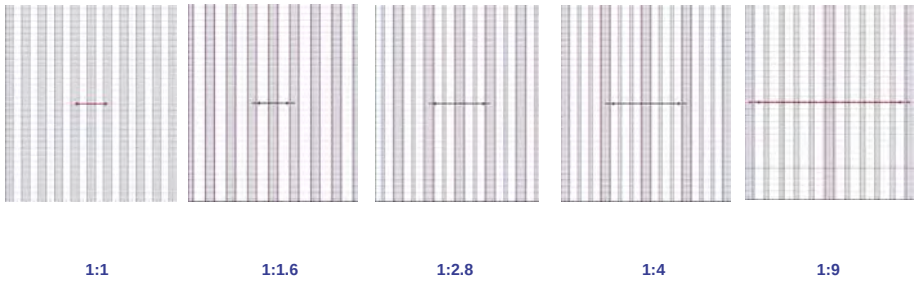
Ex. 1016, Page 363



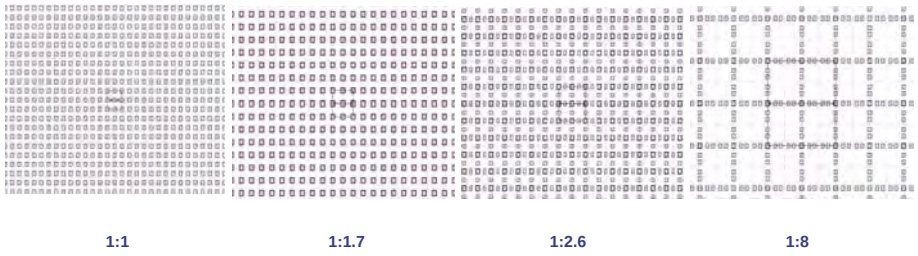
(a) Annular (b) Conventional (c) Quasar

Figure3. Illumination setting optimization result: Metal layer (a), Via layer (b, c)

Mask Design Optimizing. With help of automatic rule-based OPC optimization module in LithoCruiser^{TM}, the mask CD bias and also assisting feature rules have been optimized for improving process window for optimized illumination setting as shown in Figure 4. In Figure 4, we show assisting features between the lines and also contact holes for process window improving. The central feature both in line and contact hole arrays, is used as CD monitoring pattern.



(a) Metal layer

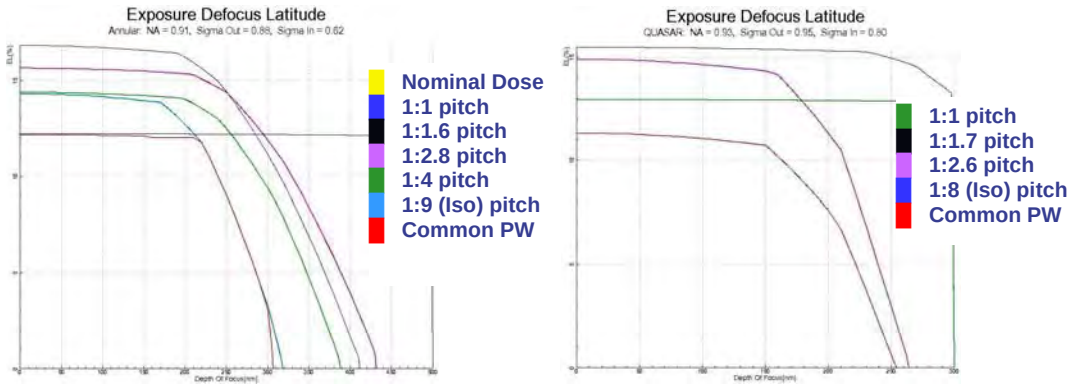


(b) Via layer

Figure3. Mask design optimizing result: Metal layer (a) and Via layer (b)

Our results show that maximum common process window for Metal layer is 12% of exposure latitude with 300nm of depth of focus at annular illumination mode and in case of Via layer, 11% of exposure latitude with 250nm of depth of focus at quasar illumination mode. The quasar illumination mode shows better process window than conventional mode. From this result, we can see that the dense feature limits the exposure latitude and isolated feature limits the depth of focus in Metal layer as shown in Figure 4(a). However, in case of Via layer, the intermediate feature limits overall process window as shown in Figure 4(b).

Ex. 1016, Page 364



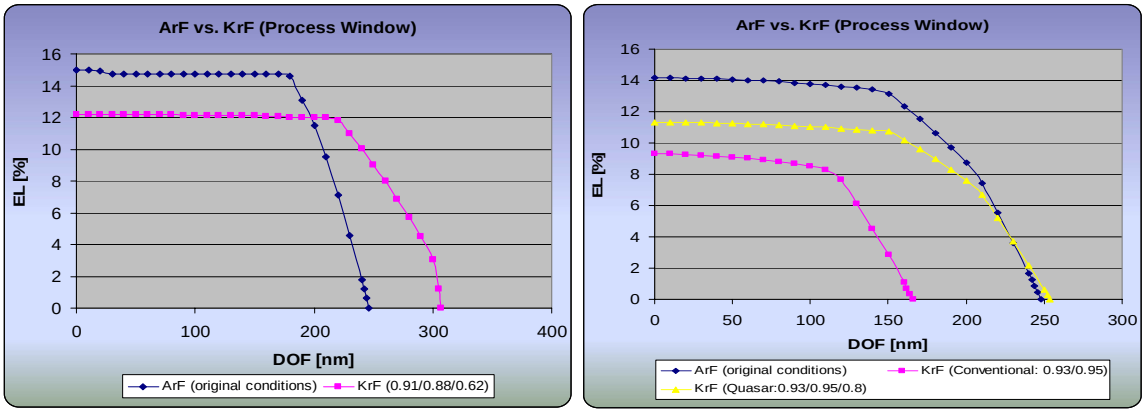
(a) Metal layer (b) Via layer

Figure 4. Optimized KrF process window: Metal layer (a) and Via layer (b)

(3) Qualification of Optimizing Result

To qualify process transfer feasibility, a comparison between ArF and KrF overlapping process windows has been done as shown in Figure 5. With respect to original process window for ArF process the one for KrF in terms of maximum exposure latitude both for Metal and Via layer decreased by about 2 ~ 3%, but it is still acceptable for production because it remains above 10%. The maximum depth of focus of KrF process for Metal layer is better than one for the reference ArF process and also, it is comparable to original one for Via layer. In case of Via layer, the quasar illumination mode shown better process window than conventional mode in terms of both exposure latitude and depth of focus.

Finally, the overall overlapping process windows show that for both Metal and Via layer the process transfer from ArF to KrF by optimization of illumination setting and mask design is feasible.



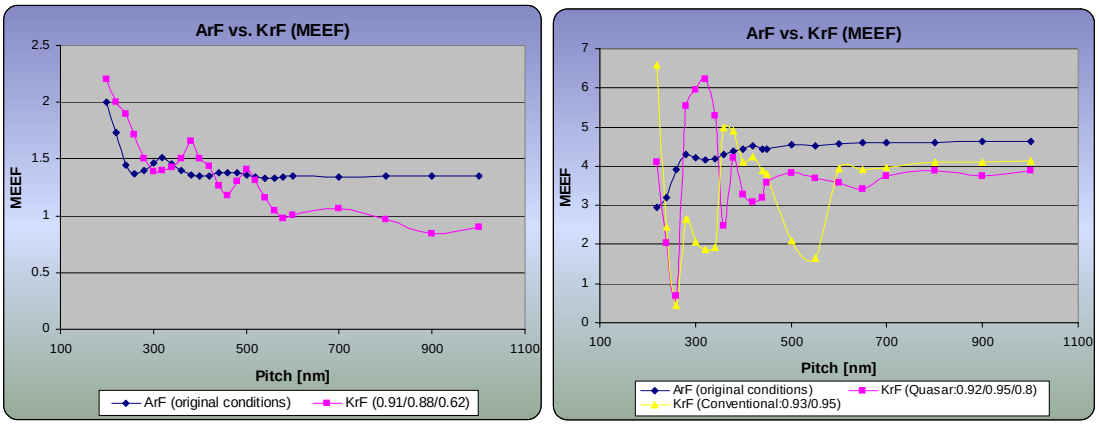
(a) Metal layer (b) Via layer

Figure 5. Common process window comparing ArF process to KrF process Metal layer (a) and Via layer (b)

MEEF Discussion

Customer CD uniformity control is one of the key items to be discussed with respect to process transfer feasibility. MEEF value through pitch for two above mentioned layers has been investigated and shown in Figure 6.In case of Metal layer, the maximum MEEF value of KrF process shows value of 2.2 and it is similar to ArF process. However, in case of Via layer, the overall MEEF value in case of KrF process with quasar illumination mode is of 1.5 times higher than one for ArF process especially within 300nm ~ 350nm pitch region. The conventional mode shows better MEEF than the quasar illumination mode except the minimum pitch. However, the process window is not bigger than with the quasar illumination mode as we showed before. In addition, the pitch for forbidden pitch region is too small to improve MEEF by adding assist feature.

As a result, the recommendation for design layout modification is to apply assist features have been given, but it is not easy to apply this solution for the mass production[1]. Another suggestion, namely, to use customized illumination mode such as soft-QUASAR™ is more efficient for improving MEEF value[2-4]. The optimization of imaging performance with soft-QUASAR™ is beyond the scope of this paper and will not be discussed, but it is clearly an option for customer to follow.



(a) Metal layer (b) Via layer

Figure 6. MEEF comparing ArF to KrF: Metal layer (a) and Via layer (b)

Conclusion

A new high NA KrF exposure tool (TWINSCAN XT:1000H, Max NA 0.93) has been released. We investigated the feasibility of process transfer from XT:1400 to XT:1000H for Metal 1 (100 nm half-pitch) and Via (110 nm half-pitch) layers of 65 nm customer product. In this paper, we suggested a 3 step process transfer approach to prove the process transfer feasibility via simulation. The transferred Metal layer shows better process window and also adequate MEEF value with respect to the original using ArF. Therefore, the Metal layer process is transferable. For Via layer the overlapping process window is comparable with the reference one with ArF, but the MEEF through pitch distribution has forbidden pitch region as a result of the boundary condition use of only standard illumination modes. Thus, to obtain better wafer CD uniformity control it is recommended to apply customized illumination mode such as soft-QUASAR™.

References

1. Joonsoo, Park et.al, *Proceedings of SPIE,* **4346**, p256 (2001)
2. Francois Weisbuch1 et. al, *Proceedings of SPIE,* **5754**, p1405 (2005)
3. Jo Finders et. al, *Proceedings of SPIE,* **6924** (2008)
4. Frank Bornebroek et. al. *Proceedings of SPIE*, **7274**, p152 (2009)

ECS Transactions, 18 (1) 341-350 (2009)
10.1149/1.3096470 © The Electrochemical Society

Process Window Analysis for KrF Photoresists under Sub-0.11um Groundrules

Lei Wang, Liang Ji, Peng Wu, Huan Kan, Weimin He, Zhiguo Zhu, Fucheng Chen, Fang Wei, Jun Zhu

Shanghai Huahong NEC Electronics Company, Ltd.
No. 1188 Chuanqiao Road, Shanghai 201206 China

Amid continuing reduction in average wafer sale price, in the past 30 years, the industry witnesses continuous shrink in production design rules in order to gain performance per unit area on each silicon wafer. While the mainstream wafer production is at 65 and 45 nm with 300 mm diameter wafers, to maintain cost competitiveness, 200 mm wafer companies also start to explore production feasibility under groundrules smaller than 110 nm while maintain the cost advantages in KrF exposure tools systems. The k1 factor under 110 nm with 248 nm illumination will be below 0.35. For example, at 90 nm, at 0.80 NA, the k1 factor is 0.29, which owns the same level of complexity in optical proximity correction compared to 65 nm at 0.93 NA with 193 nm exposure tools. However, due to that the KrF photoresists are originally designed for ground rules greater than 110 nm, such as, 130 nm structures, it is relatively unknown how feasible it is to extend their performance below 110 nm. In this paper, we will show our initial study in CD process window and CD through pitch performance at 90 nm groundrule for 3 critical layers – Active, Gate poly and Contact. The wafer data in process window and optical proximity will be analyzed and compared with performance specifications of the current exposure tools.

Introduction

The continuing semiconductor technology development has been producing more and more powerful integrated chips at cheaper and cheaper prices. In order to meet the requirement for continues shrinkage of design groundrules, photolithography has played a critical role fueled by the technology development in exposure tools, materials, and methodologies.

As we know, photolithography resolution is related to quite a few of factors such as exposure wavelength, NA (Numerical Aperture) of projection lens, illumination method, mask, chemical characteristics of photoresist, and so on. Among them, exposure wavelength is an important one. For sub-0.11 um groundrules, the wavelength for critical layers used by mainstream foundries is 193 nm, which is produced by ArF excimer lasers. At the same time, 248 nm process can not meet the same level of resolution for sub-0.11um technology node requirement due to very low k1 and difficult process control (1-2).

After a few more years of process development learning by the industry and we become more experienced in the low k1 lithography at 90 and 65 nm nodes, the difficulties associated with 90 nm on KrF wavelength becomes more and more manageable. At the

Ex. 1016, Page 369

same time, with the development of higher performance exposure tool and photoresist materials, higher resolution KrF resist and higher NA exposure tools are available to meet the requirement of sub-0.11um technology node. It will help us to get similar performance with KrF photoresist in this technology node with cheaper process cost (3).

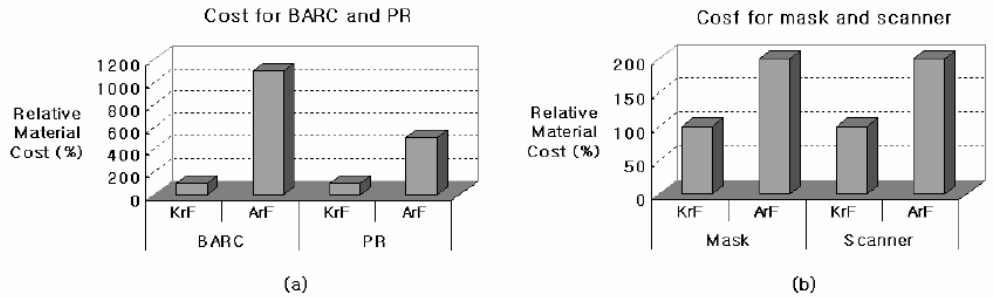


Figure 1. If KrF lithography process manufactures 90nm technology devices, the cost of fabrication can reduce. (a) cost for BARC and PR (b) cost for mask and scanner.

In this paper, we will show our study for typical critical layers – Gate poly, Metal 1, and Contact – at 90nm technology node. The process window, CD through pitch, and MEEF are checked. The result shows KrF process with high NA exposure tool is acceptable to meet the sub-0.11um technology node requirement. In the study described in the paper, the process window data are collected without scattering bars. If scattering bar is used, it is quite reasonable to believe that the overlapping process window would be even better.

Experimental

We studied Poly, Metal-1 and Contact layers in photo process. The targeted feature design rules follow a generic 90 nm logical process and a metal layer pitch 0.24 um. A KrF excimer laser scanner is used with a maximum NA of 0.82 and maximum sigma-out of 0.88 and maximum annular ratio of 0.75. For a minimum 0.24 um pitch and maximum 0.82 NA, masks are fabricated using 6% attenuated Phase Shifting Mask (PSM). Two KrF chemically amplified positive tone photoresists (one is for gate and metal, the other is for contact hole) are used.

Lithographic performance in terms of MEEF (Mask Error Enhanced Factor), DOF (Depth of Focus) and CD proximity effect is evaluated. MEEF is defined as the ratio between ADI CD vs. mask CD from nominal CD -0.01 um to target CD + 0.01 um (lx) . DOF is defined as the focal range with +/- 10% CD variation without resist thickness loss. The optical proximity effect is characterized as the CD difference for lines with pitch change from 0.22um to 1.2um under best exposure condition.

For poly layer, in order to improve the process control capability, LWR (Line Width Roughness) is evaluated as additional check item. LWR is defined as the standard deviation of line CD variation along the length direction of the line.

Ex. 1016, Page 370

Results and Discussion

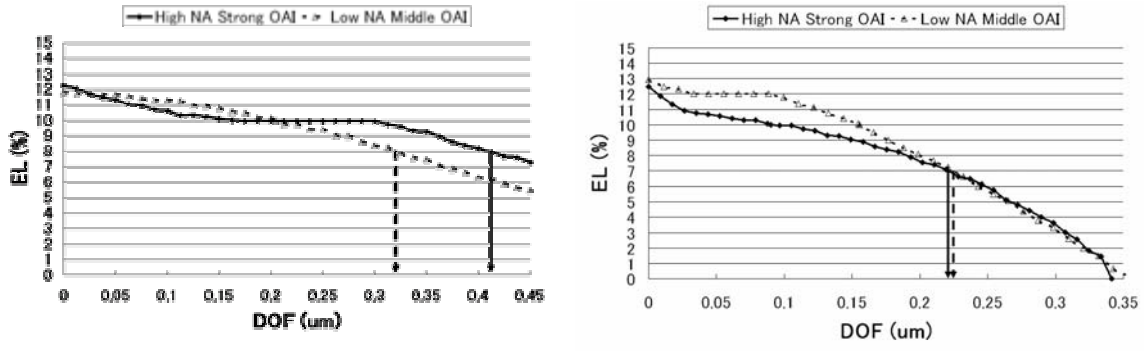
1. Poly Layer

Gate poly layer is used to define channel length of device. It is very critical to CD variation both on dense features and isolated features. Because the image contrast from a KrF process in dense features is worse than that from an ArF process, higher NA and strong off-axis illumination (OAI) are required to achieve acceptable image contrast. But too high NA and too strong OAI will cause process windows for isolated features to drop. In order to balance image contrast for dense features and process windows for isolated features, we will compare two options for this layer. One is high NA and strong OAI and the other is low NA and medium OAI.

1.1 Process Window. Displayed in Figs. 1(a) and 1(b) are process window plots from dense and isolated features, respectively. The process window data are summarized in Table I for high NA strong OAI and low NA medium OAI. The result shows that high NA and strong OAI can achieve better performance of dense features and comparable performance of isolated features. We have achieved 0.22um DOF @ 7% EL for isolated feature even without the application of scattering bars.

	Dense	ISO
High NA Strong OAI	0.41 @ 8% EL	0.22 @ 7% EL
Low NA Middle OAI	0.31 @ 8% EL	0.22 @ 7% EL

Table 1: DOF comparison between high NA strong OAI and low NA medium OAI



(a) (b)

Figure 1: (a) Dense features DOF comparison between high NA strong OAI and low NA medium OAI (b) Isolated features DOF comparison between high NA strong OAI and low NA medium OAI

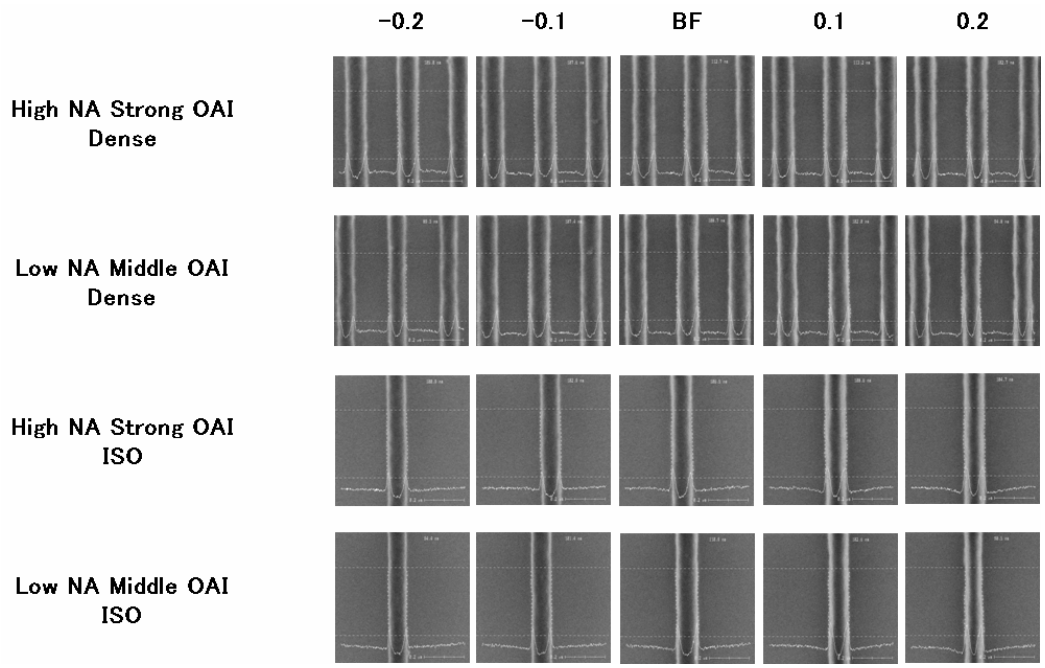
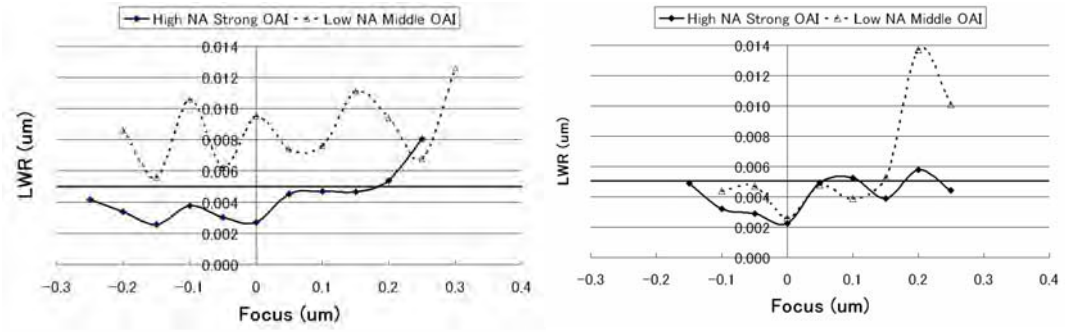


Figure 2: Top view image comparison between high NA strong OAI and low NA medium OAI depending on focus variation at poly layer



(a) (b)

Figure 3: (a) LWR comparison of dense features between high NA strong OAI and low NA medium OAI depending on focus variation at poly layer (b) LWR comparison of isolated features between high NA strong OAI and low NA medium OAI depending on focus variation at poly layer.

We have checked the dense features and isolated features CD and image comparison for focus variation at best dose condition. The result is shown in Fig. 2. Although the CD change at focus variation is comparable for the two processes, the difference in LWR is quite obvious, shown in Fig. 3. In LWR, the high NA strong OAI process is better with smoother profile.

1.2 MEEF. We have also checked the MEEF for dense features. The Fig. 3 shows the experimental ADI CD results for main features with line width from 0.10um to 0.12um. From the figure high NA strong OAI MEEF is 2.08, which is much better than that of

low NA medium OAI 2.78. It is consistent to our understanding that high NA and strong OAI could improve MEEF.

For smaller MEEF with high NA and strong OAI, we can control the CD uniformity and LWR more easily from photo process side. In future process optimization, we will fine tune the other conditions to improve MEEF.

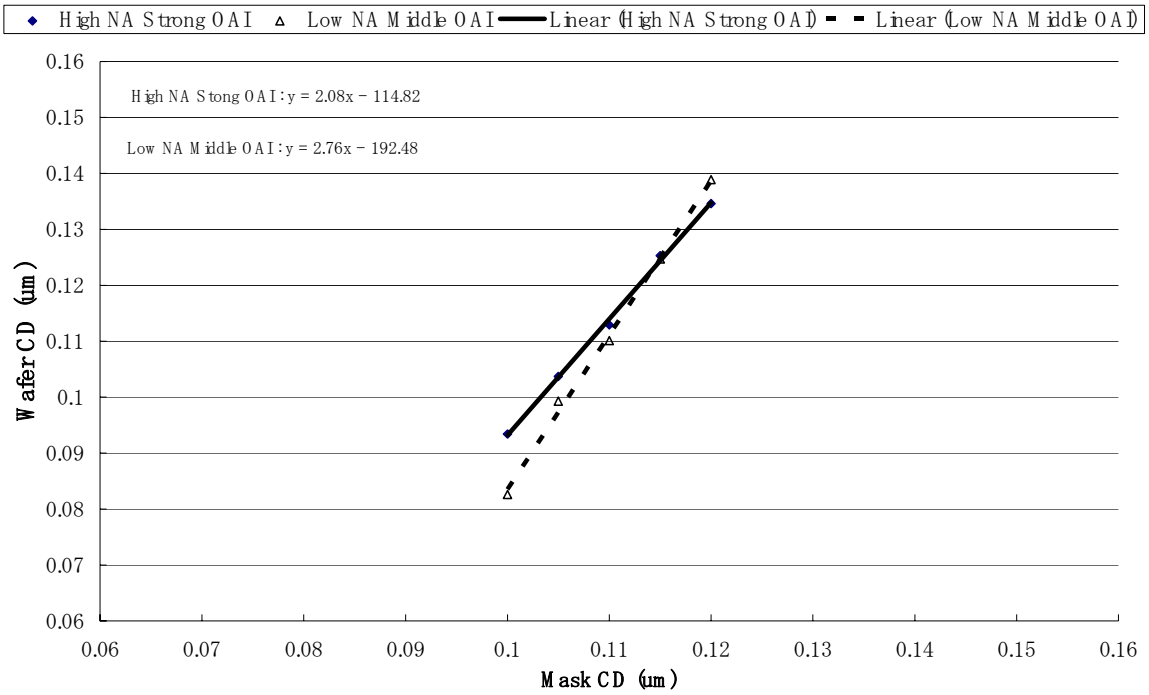


Figure 4: MEEF comparison between high NA strong OAI and low NA medium OAI

1.3 CD Proximity. The optical proximity effect is studied through CD variation with pitch. As shown in Fig. 5, the CD through pitch is almost the same for dense and semi-dense features for the two processes. The curve from high NA strong OAI has about 20 nm smaller isolated CD than that from low NA medium OAI. However, it will not gate the process. 50nm ISO-Dense bias is acceptable for photo process and OPC process. At the same time, semi-ISO and isolated feature can be improved with scattering bar if necessary.

We have also checked the through pitch LWR at the same time as we study in process window analysis. Here we don't follow the generic spec 5nm, we want to control all features LWR within 6nm and 5nm is our improve direction. The result is shown in Fig. 6. From Fig.6 we can see almost all features LWR can be controlled well for high NA strong OAI and critical features can be controlled within 5nm. It is consistent to our result in section 3.1.1.

Ex. 1016, Page 373

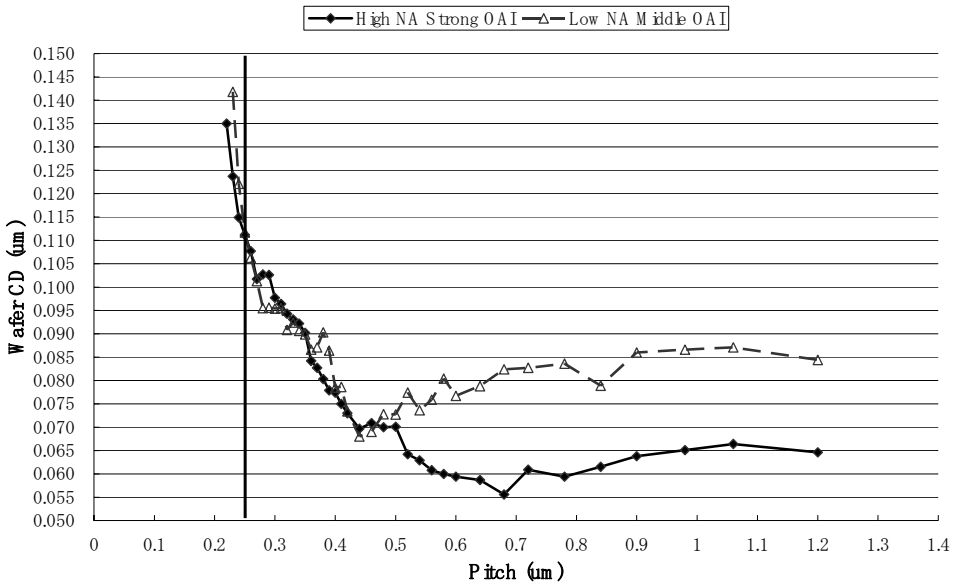


Figure 5: CD proximity comparison between high NA strong OAI and low NA medium OAI

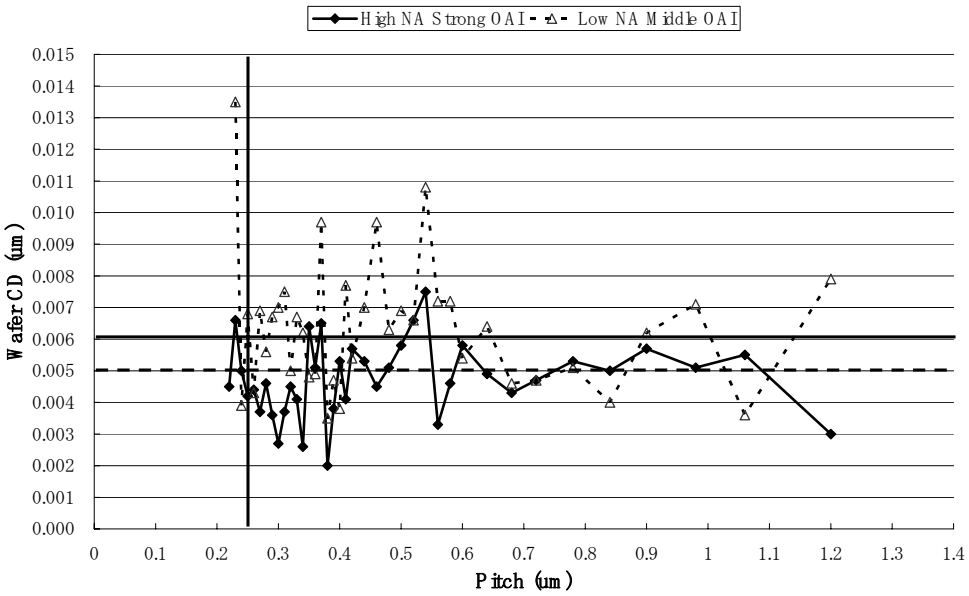


Figure 6: LWR through pitch comparison between high NA strong OAI and low NA medium OAI

With the illumination condition optimization, finally we get the acceptable overlap process window, MEEF, CD proximity and LWR. If other factors such as mask bias, bake temperature and resist evaluation can be optimized, the result could be better [4]. At the same time, scattering bar can be added to improve isolated features performance when necessary [5].

Ex. 1016, Page 374

2. Metal Layer

Different from the poly layer, metal layer is used as interconnect between different devices. The process requirement is high integrated density without bridging and shorting. As a result, the process development for metal layers will focus on tighter pitch. It needs higher resolution. In this case high NA and strong OAI is the best choice for this layer.

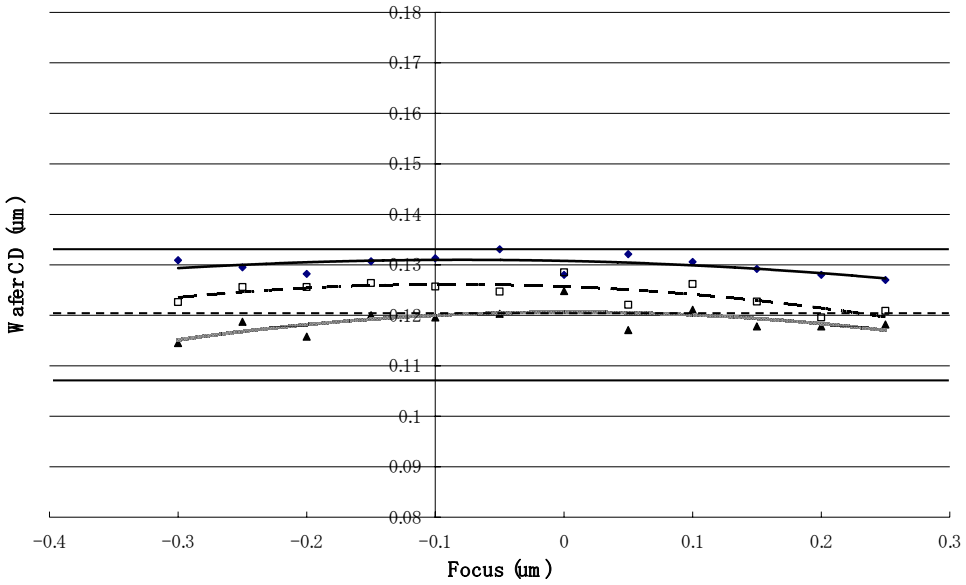


Figure 7: Process window

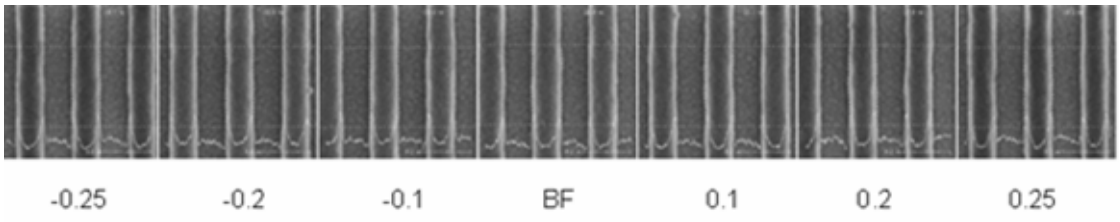


Figure 8: Top view images at various focus settings from metal layer

The process window of dense feature is shown in Fig. 7. The process window is larger enough. The DOF will be 0.55um @ 8% EL. The top view image versus focus variation is shown in Fig. 8.

The MEEF result is shown in Fig. 9. Because the limitation of thickness for the photoresist and the etch process we have used for this study and because the DARC layer SiON is not optimized, the MEEF performance is not as good as we desire, which is around 2.9. It can be improved with other module process improvement.

From these results, KrF process is acceptable for metal layer with pitch extend to 0.24um. If the ARC layer can be optimized and resist thickness can be decreased, the result will be better than we achieved.

Ex. 1016, Page 375

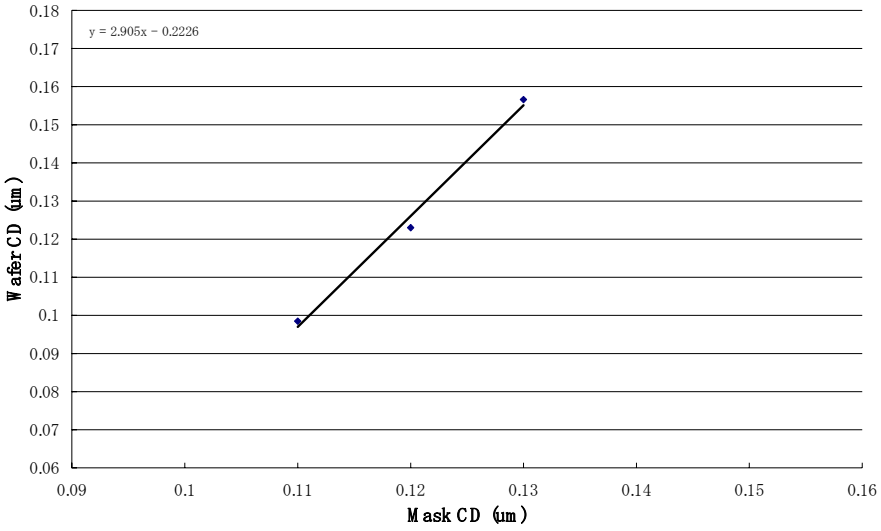


Figure 9: MEEF measured for metal layer

3. Contact Layer

Next we studied the KrF process capability for the contact hole layer. In general, it is very challenging for the contact hole or via layers when k1 becomes small. In addition, the hole layer OPC can also be very challenging when the CD proximity range is too large. In this case, high NA conventional setting is probably the best for a generic logical process since we would like to both get acceptable process window and well controlled CD proximity. The MEEF requirement is not as tight as line and space layer.

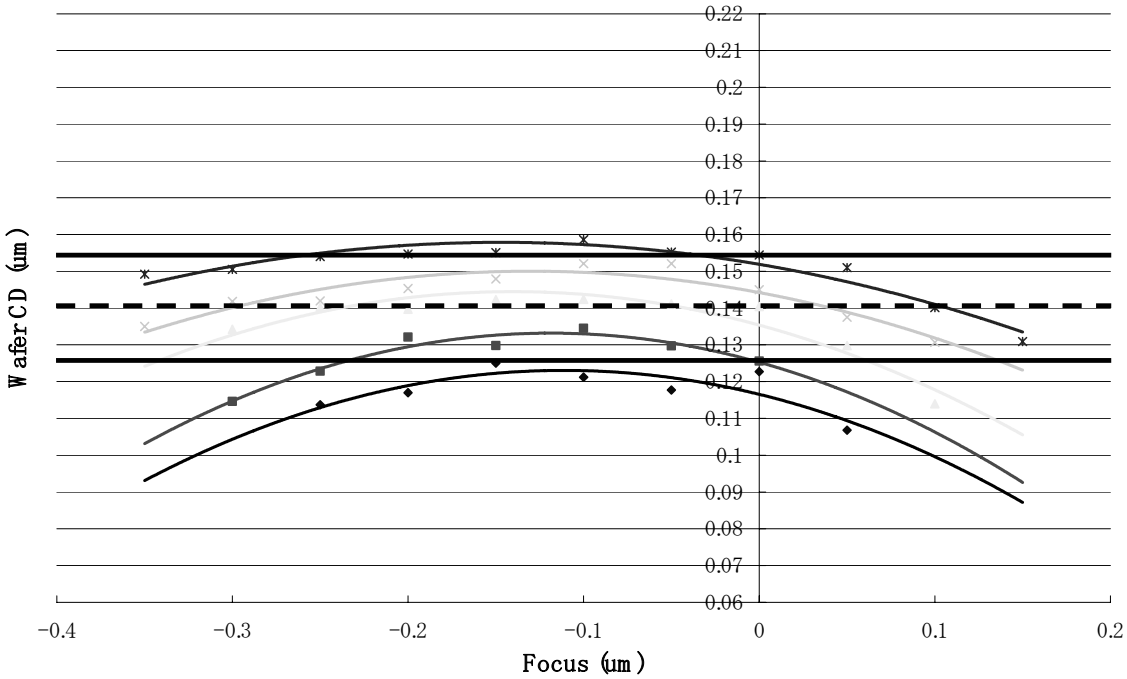


Figure 10: Process window

Ex. 1016, Page 376

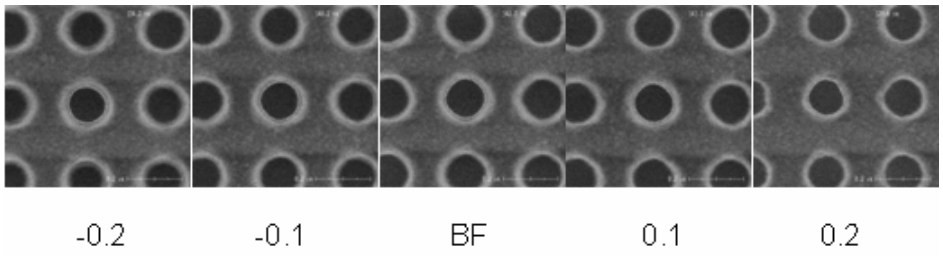


Figure 11: Top view images at various focus settings from the contact layer

The process window of dense feature is shown in Fig. 10. The DOF has been measured to be 0.26um @ 10% EL, which is acceptable. The top view image versus focus variation is shown in Fig. 11.

The optical proximity behavior is shown with CD variation with pitch same as poly layer. From Fig. 12, we can control the CD proximity to within 10nm without OPC. It is very friendly pre-condition to OPC process.

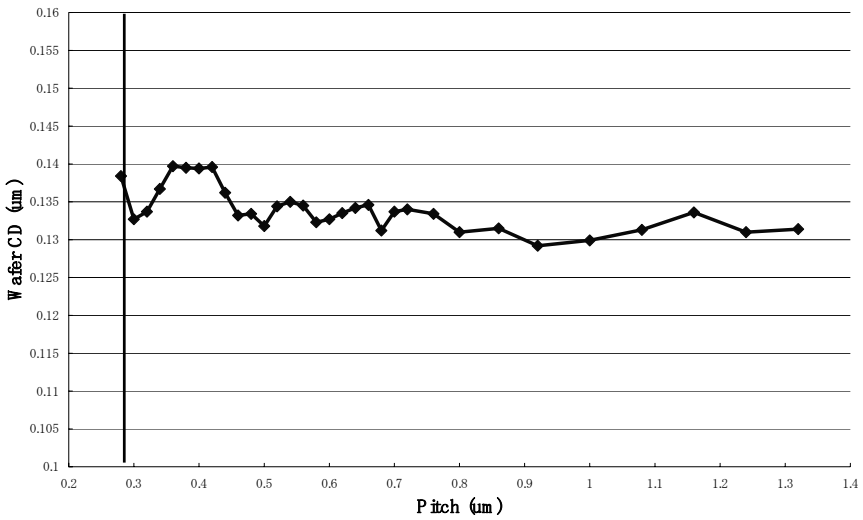


Figure 12: CD proximity for contact layer

MEEF check result is shown in Fig. 13. It is limited by exposure tool and illumination methods and the number is 4.44. Since we need to control isolated hole process window and CD proximity, so the NA is not large enough to reduce the MEEF. It is a little larger as we want. Fortunately, the MEEF for contact hole around 4.4 is actually not so bad. Due to the fact that we have already achieved good isolated hole process window, proximity and MEEF, the slightly high MEEF is acceptable.

From these results, KrF process is acceptable for contact layer.

Ex. 1016, Page 377

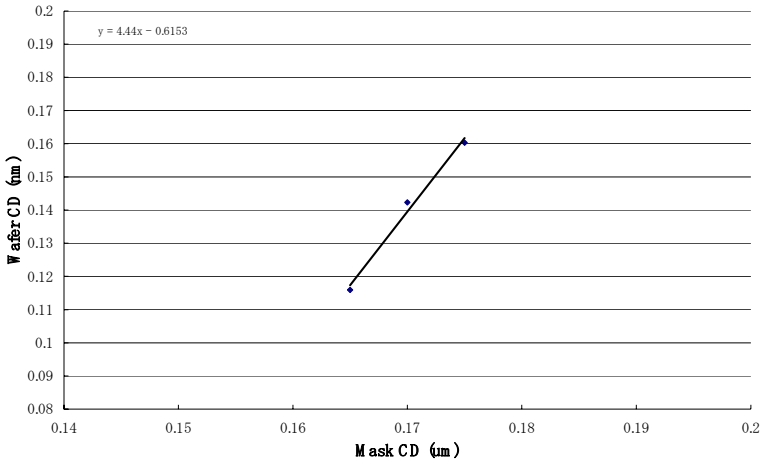


Figure 13: MEEF for the contact process

Conclusions

We have performed a study on KrF process capability for critical layers at 90nm technology node. We have found that KrF process can get enough process control capability for this generation to instead of ArF process to reduce the cost although current KrF scanners with maximum 0.82 NA can not get the similar good image contrast as those ArF scanners especially in EL and MEEF control. On the other side, we have found that KrF resist LWR performance is acceptable for 90 nm process requirement.

Acknowledgments

The author of this paper would like to thank the production division, the technology division, the third engineering division, the design service division, and the management teams of Shanghai Huahong NEC Electronics Company, limited for their kind support of the engineering work for this paper.

References

1. Shangho Lin, et. al., *How to Print 100nm Contact Hole with Low NA 193nm Lithography*, Proc. SPIE **5376,** 1091 (2004).

2. Jo Finders, et. al., C*an DUV Take Us Below 100nm*, Proc. SPIE **4346,** 153 (2001).

3. Youngdoo Jeon, et. al., *90nm node Contact Hole Patterning through applying Model Based OPC in KrF Lithography*, Proc. SPIE **6924**-103 (2008) and references therein.

4. Lei Wang, et. al., "The Characterization of Photoresist for Accurate Simulation Beyond Gaussian Diffusion", Proc. SPIE **6519-35** (2007) and references therein.

5. H.T. Line, et. al., "SubO.18 tm Line/Space Lithography using 248nm Scanners and Assisting Feature OPC Masks", Proc. SPIE **3873,** 3873 (1999).

Ex. 1016, Page 378

ECS Transactions, 18 (1) 351-356 (2009)
10.1149/1.3096471 © The Electrochemical Society

Computation lithography: virtual reality and virtual virtuality

Alfred K. K. Wong[a] and Edmund Y. Lam[b]

[a]Magma Design Automation, Inc., 1650 Technology Drive, San Jose, CA 95110, U. S. A.
[b]Department of EEE, The University of Hong Kong, Pokfulam Road, Hong Kong*

Abstract

Computation lithography is enabled by a combination of physical understanding, mathematical abstraction, and implementation simplification. An application in the virtual world of computation lithography can be a virtual reality or a virtual virtuality depending on its engineering sensible-ness and technical feasibility.

1 Introduction

The current prevalence of computation lithography traces its root to the 1970s, when it was discovered that photoresist dissolution can be modeled as a surface etching phenomena, with the etch rate dependent on the state of photochemistry rather than the manner in which the resist was exposed.[1,2] But it has taken some time before lithography simulation, the precursor of computation lithography, establishes its present-day credibility.[i] Even as late as the 1980s lithography simulation was still labeled by some as "Yesterday's technology simulated tomorrow".[5]

Nevertheless, the 1970s and 1980s was a germinating period for lithography simulation; many fundamental algorithms were developed. Computation of partially coherent imaging was placed on a firm foundation,[6] permitting accurate simulation of projection printing in the regime where neither the relationship between mask and intensity nor that between objects and image-plane field is linear. A variety of techniques, including the waveguide method,[7,8] and finite-element[9,10] and finite-difference approaches,[11,12] were developed to examine scattering and diffraction of light propagation through mask openings. Resist development modeling, which had been plagued by looping instability in which surfaces of dissolved materials curve back onto itself,[13] was properly handled by the fast-marching level-set method.[13,14]

These algorithms enhanced a lithographer's arsenal of tools, complementing empirical data supplied by wafer exposures. With exponential improvement in computer speed and rising equipment cost, simulation imperceptibly assumed an ever-increasing role in lithography process development. During the 1970s and 1980s, however, algorithm performance still restricted analysis to dimensions on the order of a feature size.

Adoption of optical proximity correction (OPC) in the 1990s triggered further innovation. Success of rule-based OPC[15,16] quickly translated to a demand for model-based OPC,[17–20] which clamored full-chip image computation capabilities. Coherent decomposition[21–23] was equal to the challenge. By approximating a partially coherent optical system as a weighted collection of coherent systems, computation complexity was reduced by orders of magnitude.

2 Ingredients of computation lithography

Model-based OPC is an epitome of computation lithography, which can be defined as the use of computer algorithms in lithography-related integrated circuit (IC) development. The basic components of a computation lithography application are

*The work described in this paper was partially supported by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Numbers 713906 and 713408).

[i]The history of lithography simulation is considered from a few points of view in Dill,[3] Mack,[4] and Neureuther.[5]

Ex. 1016, Page 379

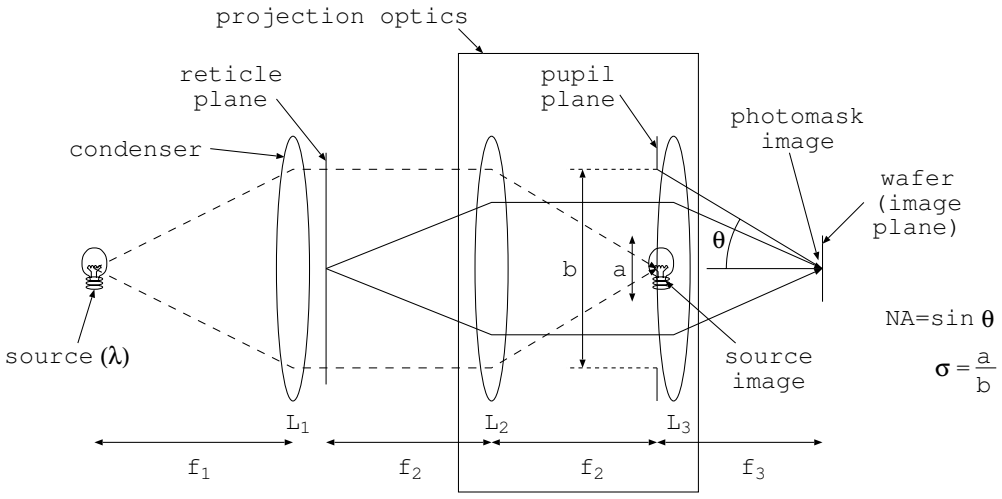


Figure 1: Illustration of a projection microlithography exposure system.[26]

1. specifications;
2. physical understanding;
3. mathematical abstraction; and
4. implementation simplification.

Let us illustrate these aspects using full-chip image computation as an example.

2.1 Specifications

The goal of full-chip image computation is to calculate the image intensities at select locations (or critical dimensions of select features) on a design quickly and accurately. Depending on the application, "quickly" and "accurately" are specified by concrete numbers such as 1 s per intensity value and 1 nm maximum deviation from empirical calibration data.

2.2 Physical understanding

In order to compute an image, a comprehension of image formulation principles is necessary. From an understanding of light diffraction, the physics of a projection microlithography system can be distilled, as illustrated in Fig. 1.[24,25] A quasi-monochromatic light source of mean wavelength λ illuminates the reticle via the condenser lens L_1. In conjunction with L_1, the lens L_2 forms an image of the source in the pupil. This lens L_2 performs a Fourier transform of the object such that the field across the pupil plane is the spectrum of the object. Low-spatial-frequency components pass closer to the center of the pupil, and higher-frequency components are nearer to the periphery of the pupil. The highest frequencies are cut off by the pupil. The lens L_3 forms the image by combining the filtered frequency components.

2.3 Mathematical abstraction

Before an algorithm for image computation can be formulated, operation of the exposure system depicted in Fig. 1 needs to be described mathematically. Image intensity at a point (x, y) is given by:[27]

$$I(x,y) = \int_{-\infty}^{+\infty}\!\!\!\!\iiint w(x-x'_o, y-y'_o; x-x''_o, y-y''_o) O(x'_o, y'_o) O^*(x''_o, y''_o)\, dx'_o dy'_o dx''_o dy''_o, \tag{1}$$

where

$$\begin{aligned} w(x'_o, y'_o; x''_o, y''_o) &= J(x'_o - x''_o, y'_o - y''_o) H(x'_o, y'_o) H^*(x''_o, y''_o), \\ J(x'_o - x''_o, y'_o - y''_o) &\quad \text{is the mutual intensity,} \\ H(x, y) &\quad \text{is the optical system transfer function,} \\ O(x, y) &\quad \text{is the mask transmittance,} \end{aligned}$$

and the asterisk $*$ represents complex conjugation. Equation (1) says that the image intensity at (x,y) can be interpreted as the weighted sum (w) of contributions from neighboring objects $O(x'_o, y'_o)$ and $O(x''_o, y''_o)$.

2.4 Implementation simplification

Computation complexity of Eq. (1) is high because of the four-fold integral. To improve efficiency, the function w can be decomposed into a series of its eigenvectors φ_k and corresponding eigenvalues λ_k,[21] such that

$$w(x'_o, y'_o; x''_o, y''_o) = \sum_{k=1}^{\infty} \lambda_k \varphi_k(x'_o, y'_o) \varphi_k^*(x''_o, y''_o).$$

For typical exposure systems, the magnitudes of a few eigenvalues are much larger than the rest. We can thus approximate the image as a finite sum of, say, K dominant eigenvectors:

$$I(x, y) \approxeq \sum_{k=1}^{K} \lambda_k \left| \iint_{-\infty}^{+\infty} \varphi_k(x - x', y - y') O(x', y') \, dx' dy' \right|^2. \tag{2}$$

Although Eq. (2) reduces the four-fold integral of Eq. (1) to K two-dimensional convolutions, computation complexity is still high. Nevertheless, for layouts that can be decomposed exclusively into rectangles, we can represent each rectangle as the sum of four quadrant functions

$$Q(x, y) = \begin{cases} 1 & \text{if } x \geq 0 \text{ and } y \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

such that

$$\begin{aligned} O(x, y) = \sum_{n=1}^{N} t_{\text{fg}}^{(n)} \big[& Q(x - x_1^{(n)}, y - y_1^{(n)}) - Q(x - x_2^{(n)}, y - y_1^{(n)}) + \\ & Q(x - x_2^{(n)}, y - y_2^{(n)}) - Q(x - x_1^{(n)}, y - y_2^{(n)}) \big], \end{aligned}$$

where N is the total number of rectangles, the superscript (n) indexes the rectangular shapes, and $t_{\text{fg}}^{(n)}$ represents the transmittance of the nth shape.[22,23] Substituting the above expression into Eq. (2) results in

$$\begin{aligned} I(x, y) \approxeq \sum_{k=1}^{K} \left| \sum_{n=1}^{N} \big[& \psi_k(x - x_1^{(n)}, y - y_1^{(n)}) - \psi_k(x - x_2^{(n)}, y - y_1^{(n)}) + \right. \\ & \left. \psi_k(x - x_2^{(n)}, y - y_2^{(n)}) - \psi_k(x - x_1^{(n)}, y - y_2^{(n)}) \big] \right|^2, \end{aligned} \tag{3}$$

where

$$\psi_k(x, y) = \sqrt{\lambda_k} \cdot Q(x, y) \otimes \varphi_k(x, y)$$

is the convolution of the quadrant function with the kth eigenvector scaled by the eigenvalue. Since the functions $\psi_k(x, y)$ are independent of the object, they can be precomputed and stored. Image calculation hence becomes an addition operation, improving computation efficiency tremendously.

Ex. 1016, Page 381

3 Virtual reality and virtual virtuality

In the virtual world of computation lithography, virtual realities are those applications that perform useful service for IC development. They are tools characterized by

1. specifications that make engineering sense, and
2. performance (based on physical understanding, mathematical abstraction, and implementation simplification) that meets the specification targets.

Virtual virtualities, on the other hand, are algorithms that fall short of one or both of the two criteria. That model-based OPC and post-OPC verification[28,29] are virtual realities there is little doubt; they are incumbents in a standard production flow. But it is of interest to examine the degree of reality of a few applications at the fringe.

3.1 Inverse lithography

Inverse lithography aims to derive the optimal mask given a desired image.[30,31] Since the synthesized mask is unconstrained by the topology of the original physical design, inverse lithography promises to deliver superior performance over existing model-based OPC approaches. The specifications of inverse lithography make engineering sense, for they are essentially those of OPC's. However, current performance still falls short, primarily because of the lack of a mathematical description for mask regularization[32] that provides a satisfactory balance between lithography robustness and mask manufacturability.

In a similar manner, determining printability of a layout using a generic model[33] is hindered by deficiency of the mask model, and concurrent source-mask optimization[34] needs to resolve both mask and source regularization issues.

3.2 Physical DFM

Physical design-for-manufacturability (DFM) refers to remedying weaknesses in a physical design.[35] It comprises finding hotspots by model-based verification, and fixing them by polygon movements. Since these two steps are rather straightforward, there is little implementation difficulty. The issue lies with specifications.

During the hotspot analysis stage in physical DFM, it is sometimes impractical, for turn-around or for accessibility reasons, to apply the same algorithm as the one used for sign-off verification. Differences between the DFM and the sign-off algorithms would result in hotspot discrepancy, leading to imperfect accuracy such as misses (hotspots not being remedied) and extraneous fixing (repairing of non-hotspots). The achievable accuracy depends on algorithm precision. For example, with a critical dimension simulation variability of 5% ($1\,\sigma$), the best average matching and extra rates are 73.6% and 26.4% respectively.[36] More stringent accuracy specifications than such theoretical bounds are unrealistic.

3.3 Electrical DFM

From a full-chip image lithography contours can be derived. By employing a suitable wire delay or transistor model such as BSIM,[37] electrical performance can be calculated and fed back for OPC or layout enhancement. Such electrical DFM approach was shown to be viable from the computational standpoint.[38] Specifications for these applications, however, must consider all factors that significantly affect electrical behavior, an example being the possible interaction between corner-rounding effects and oxidation-induced stress effects. Should stress effects be found to be of higher relative importance, contour-to-electrical DFM may not make full engineering sense.

4 Remarks

If we consider a plane with the x-axis denoting the degree of engineering sensible-ness and the y-axis representing performance requirement (defined by some combination of turn-around, accuracy, and precision

Ex. 1016, Page 382

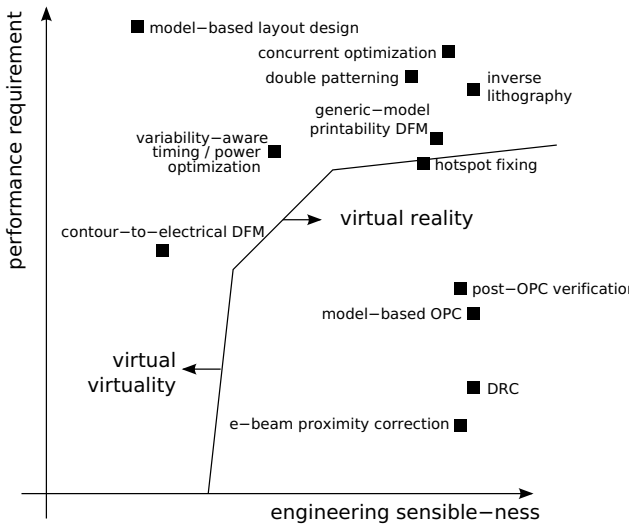


Figure 2: The reality-virtuality space of computation lithography (not to scale).

metrics), we can demarcate the plane into two areas. The one to the lower-right of the dividing line is the realm of virtual reality, while the region to the upper-left is the region of virtual virtuality. Applications can be placed onto this plane according to their degrees of reality, as illustrated in Fig. 2.

With algorithmic innovation and increased parallelization, the boundary between reality and virtuality will evolve. Today's virtuality can become tomorrow's reality. Simultaneously, with enhancements in our understanding and as requirements scale with technology, the markers also travel on their own trajectories. For example, lithography model accuracy needs to improve with decreasing critical dimension, pushing the "model-based OPC" marker upwards in the graph. Today's reality may become tomorrow's virtuality.

For computation lithographers, the goals are to enlarge the virtual reality region and to push the markers towards the lower-right. In this regard, applications on the right side of Fig. 2 around the reality envelope deserve special attention.

References

[1] F. H. Dill, W. Hornberger, P. Hauge, and J. Shaw, "Characterization of positive photoresists," *IEEE Transactions on Electron Devices*, vol. ED-22, pp. 445–452, July 1975.

[2] K. L. Konnerth and F. H. Dill, "In-situ measurement of dielectric thickness during etching or developing process," *IEEE Transactions on Electron Devices*, vol. ED-22, pp. 452–456, July 1975.

[3] F. Dill, "The basis for lithographic modeling," in *Proc. SPIE* (B. W. Smith, ed.), vol. 5754, pp. 377–382, 2005.

[4] C. A. Mack, "Thirty years of lithography simulation," in *Proc. SPIE* (B. W. Smith, ed.), vol. 5754, pp. 1–12, 2005.

[5] A. Neureuther, "If it moves, simulate it!," in *Proc. SPIE* (H. J. Levinson and M. V. Dusa, eds.), vol. 6924, p. 692402, 2008.

[6] M. Yeung, "Modeling aerial images in two and three dimensions," in *Proc. Kodak Microelectronics*, pp. 115–126, 1985.

[7] D. Nyyssonen and C. P. Kirk, "Optical microscope imaging of lines patterned in thick layers with variable edge geometry: theory," *J. Opt. Soc. Am. A*, vol. 5, pp. 1270–1280, Aug. 1988.

[8] K. Lucas, C.-M. Yuan, and A. Strojwas, "A rigorous and practical vector model for phase shifting masks in optical lithography," in *Proc. SPIE* (J. D. Cuthbert, ed.), vol. 1674, pp. 252–263, 1992.

[9] T. Matsuzawa, A. Moniwa, N. Hasegawa, and H. Sunami, "Two-dimensional simulation of photolithography on reflective stepped substrate," *IEEE Transactions on Computer-aided Design of Integrated Circuits and Systems*, vol. 6, pp. 446–451, May 1987.

[10] H. P. Urbach and D. A. Bernard, "Modeling latent image formation in photolithography using the Helmholtz equation," in *Proc. SPIE* (V. Pol, ed.), vol. 1264, pp. 278–293, 1990.

[11] K. S. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," *IEEE Transactions on Antennas Propagation*, vol. 14, pp. 302–307, May 1966.

[12] R. Guerrieri, K. H. Tadros, J. Gamelin, and A. Neureuther, "Massively parallel algorithms for scattering in optical lithography," *IEEE Transactions on Computer-aided Design of Integrated Circuits and Systems*, vol. 10, pp. 1091–1100, Sept. 1991.

Ex. 1016, Page 383

[13] J. A. Sethian, "Fast marching level set methods for three-dimensional photolithography development," in *Proc. SPIE* (G. Fuller, ed.), vol. 2726, pp. 262–272, 1996.

[14] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics*, vol. 79, pp. 12–49, 1988.

[15] J. F. Chen, T. Laidig, K. Wampler, and R. Caldwell, "Optical proximity correction for intermediate-pitch features using sub-resolution scattering bars," *J. Vac. Sci. Technol. B*, vol. 15, pp. 2426–2433, Nov. 1997.

[16] O. Otto and R. Henderson, "Advances in process matching for rules-based optical proximity correction," in *Proc. SPIE*, vol. 2884, pp. 425–434, 1996.

[17] N. Cobb and A. Zakhor, "Experimental results on optical proximity correction with variable threshold resist model," in *Proc. SPIE* (G. Fuller, ed.), vol. 3051, pp. 458–468, 1997.

[18] M. Rieger and J. Stirniman, "Mask fabrication rules for proximity corrected patterns," in *Proc. SPIE*, vol. 2884, pp. 323–332, 1996.

[19] T. Waas, H. Eisenmann, and H. Hartmann, "Proximity correction for high CD-accuracy and process tolerance," in *Proc. Symposium on Nanocircuit Engineering*, 1994.

[20] H.-Y. Liu, L. Karklin, Y.-T. Wang, and Y. C. Pati, "Application of alternating phase-shifting masks to 140 nm gate patterning II: Mask design and manufacturing tolerances," in *Proc. SPIE*, vol. 3334, pp. 2–14, 1998.

[21] H. Gamo, "Matrix treatment of partial coherence," in *Progress in Optics* (E. Wolf, ed.), vol. 3, pp. 187–332, North-Holland, 1964.

[22] Y. C. Pati, A. A. Ghazanfarian, and R. F. Pease, "Exploiting structure in fast aerial image computation for integrated circuit patterns," *IEEE Transactions on Semiconductor Manufacturing*, vol. 10, pp. 62–74, Feb. 1997.

[23] A. E. Rosenbluth, G. Gallatin, R. Gordon, W. Hinsberg, J. Hoffnagle, F. Houle, K. Lai, A. Lvov, M. Sanchez, and N. Seong, "Fast calculation of images for high numerical aperture lithography," in *Proc. SPIE* (B. Smith, ed.), vol. 5377, pp. 615–628, 2004.

[24] M. Born and E. Wolf, *Principles of Optics*, pp. 491–555. Pergamon Press, sixth ed., 1980.

[25] J. W. Goodman, *Introduction to Fourier Optics*. Roberts and Company, third ed., 2004.

[26] A. K. Wong, *Resolution Enhancement Techniques in Optical Lithography*. SPIE Press, 2001.

[27] A. K. Wong, *Optical Imaging in Projection Microlithography*. SPIE Press, 2005.

[28] J. Kim and M. Fan, "Hotspot detection on post-OPC layout using full-chip simulation-based verification tool: a case study with aerial image simulation," in *Proc. SPIE* (K. R. Kimmel and W. Staud, eds.), vol. 5256, pp. 919–925, Dec. 2003.

[29] S. D. Shang, Y. Granik, N. B. Cobb, W. Maurer, Y. Cui, L. W. Liebmann, J. M. Oberschmidt, R. N. Singh, and B. R. Vampatella, "Failure prediction across process window for robust OPC," in *Proc. SPIE* (A. Yen, ed.), vol. 5040, pp. 431–440, June 2003.

[30] L. Pang, Y. Liu, and D. Abrams, "Inverse lithography technology (ILT): What is the impact to the photomask industry?," in *Proc. SPIE* (M. Hoga, ed.), vol. 6283, p. 62830X, Apr. 2006.

[31] S. H. Chan, A. K. Wong, and E. Y. Lam, "Initialization for robust inverse synthesis of phase-shifting masks in optical projection lithography," *Optical Express*, vol. 16, pp. 14746–14760, Sept. 2008.

[32] A. Poonawala and P. Milanfar, "Mask design for optical microlithography—an inverse imaging problem," *IEEE Transactions on Image Processing*, vol. 16, pp. 774–788, Mar. 2007.

[33] B. Yenikaya and A. Sezginer, "A rigorous method to determine printability of a target layout," in *Proc. SPIE* (A. K. K. Wong and V. K. Singh, eds.), vol. 6521, p. 652112, Mar. 2007.

[34] A. E. Rosenbluth, S. Bukofsky, C. Fonseca, M. Hibbs, K. Lai, A. F. Molless, R. N. Singh, and A. K. K. Wong, "Optimum mask and source patterns to print a given shape," *Journal of Microlithography, Microfabrication, and Microsystems*, vol. 1, pp. 13–30, Apr. 2002.

[35] H. Mashita, T. Kotani, F. Nakajima, H. Mukai, K. Sato, S. Tanaka, K. Hashimoto, and S. Inoue, "Tool-induced hotspot fixing flow for high volume products," in *Proc. SPIE*, vol. 7028, p. 70283I, 2008.

[36] A. K. K. Wong and E. Y. Lam, "The nebulous hotspot," in *Proc. SPIE*, vol. 7275, 2009.

[37] J. H. Huang, Z. H. Lui, M. C. Jeng, P. K. Ko, and C. Hu, "A robust physical and predictive model for deep-submicrometer MOS circuit simulation," Master's thesis, University of California, Berkeley, May 1993. Memorandum No. UCB/ERL M93/57.

[38] S. Banerjee, P. Elakkumanan, L. W. Liebmann, J. A. Culp, and M. Orshansky, "Electrically driven optical proximity correction," in *Proc. SPIE* (V. K. Singh and M. L. Rieger, eds.), vol. 6925, p. 69251W, 2008.

ECS Transactions, 18 (1) 357-366 (2009)
10.1149/1.3096472 © The Electrochemical Society

OPC Accuracy and Development Time Improvement for Advanced Technology Nodes

Peter Nikolsky, Tjitte Nooitgedagt, Paul van Adrichem, Jeroen Meessen, Carsten Kohler, Baukje Wisse[a], Stefan Hunsche[b]
[a] ASML Netherlands B.V., De Run 6501, 5504DR Veldhoven The Netherlands;
[b] Brion, an ASML company, 4211 Burton Drive, Santa Clara, CA 95054, USA

Double Dipole (DDL), Double Patterning (DP), Spacers and other advanced process technologies require an enormous amount of CD and tool data collection and development time for OPC modeling.

In this paper we studied a way to reduce the overall OPC preparation time and increase the OPC accuracy by early and accurate collection of the OPC measurements, automatic CD-SEM recipe generation and CD data analysis and model calibration with Brion's Tachyon software.

We reviewed the design of the OPC test chip, OPC exposures, data analysis and OPC modeling on the advanced and accurate Tachyon T2.0 OPC platform and ways to improve modeling cycle time.

Another critical parameter of the OPC modeling is accuracy as one of the main parts of the CD error budget. We investigated approaches to OPC modeling accuracy and achieved RMS below 1nm for critical features by using module data of the respective ASML TWINSCAN XT:1900i scanner.

Introduction

Aggressive shrink of design rules to 45nm and below dramatically increases the amount of CD-SEM measurements, data preparation, calculations required for OPC modeling demand high CD prediction accuracy through different focus and expose conditions.

OPC now faces two main issues: hardly acceptable cycle time for calibration and limited accuracy achieved by "traditional" parametric OPC based on limited use of measured tool and process parameters.

The resulting fit and calibration cycle time depends on the quality and amount of the input CD data and measured tool parameters and model quality. Unfortunately these two issues are antagonistic: improving OPC accuracy needs an increase OPC cycle time because more CD data and tool parameters have to be measured and better modeling has to be done.

The goal of our study is achieve OPC modeling accuracy required for the technology node 45nm and below in a two-three weeks timeframe at the early development stage required for time-to-market reduction.

The key points of our approach are:

- Early access to the advanced ASML TWINSCAN XT:1900i scanner for tool parameters collection and exposures at the ASML Demo Lab in Veldhoven;

Ex. 1016, Page 385

- Use Tachyon OPC software and hardware for accurate and fast Litho modeling;
- Precise CD-SEM measurements and CD data processing;
- Optimized calibration flow to achieve better accuracy in a short runtime.

During this feasibility study we have achieved the following results:

- The shortest development time for the OPC model;
- High simulation accuracy with the RMS below 1nm for the 1D features at the nominal Dose and Focus conditions and below 3nm through Process Window;
- Early start of the initial OPC modeling leading to reduced cycle time of the whole Litho module;
- TWINSCAN XT:1900i scanner parameters were ranked in respect to the influence on the OPC model fit to optimize data collection and calibration process.

Design node and target feature for the calibration

The chosen Design node for calibration is 45nm. The target feature is the Dense Line with 90nm pitch (L:S=1:1). This feature choice mimics a Flash-like device requiring the highest available OPC accuracy for the densest features on the chip.

We use two different alt-PSM test reticles developed at ASML: one designed for the 45nm process (Proxi45 mask) and another one containing features for various design nodes from 32nm to 250nm (low-k1 test mask).

Search for the best illumination settings and resist process choice for the OPC calibration

The Illumination settings were chosen according to an experiment performed to find the most suitable scanner setting to print 45nm lines, 65nm spaces and 80nm line gaps through pitch with sufficient process latitude.

Calibrating the OPC model with high accuracy (low rms residue) requires good image fidelity and small variation of a given feature (gauge) at the nominal dose and focus setting. This in turn requires sufficient image contrast and focus latitude to minimize scanner induced CD variation across the field and across the wafer which is used for the calibration measurements. To resolve the aggressive pitch of 90nm at 1.35NA off-axis illumination is required. The sigma center value and pole width affect the trade-off between contrast at the minimum pitch and focus latitude through the entire pitch range. Halftone masks (6% AttPSM) are used.

We considered pitches from 90nm to 1000nm and various Annular and C-Quad illumination settings.

Assist feature (AF) rules were applied for different pitch ranges of lines to reduce the focus sensitivity to an acceptable level.

Space and line gap printing was analyzed without the use of AF. Line gaps were chosen to mimic 2D features (As a matter of fact they are only partly isolated in the direction along the line. i.e. the gap size is small. Increasing the gap size leads to large focus sensitivity and would require complex AFs).

Ex. 1016, Page 386

All metrics (CD, EL at 25nm defocus and MEEF) have been calculated through pitch. Different doses have been considered which effects the mask biases to print on target (AFs were not biased).

Following the simulation results the C-Quad illumination with 20 deg. poles and Sigma 0.9/0.65 was chosen.

The preliminary test to verify that the ASML in-house resist process is suited for 32nm node model calibration (45nm lines, 90nm pitch and up and line ends) was performed before the calibration.

The process conditions of the experiment were:

- ASML in-house resist process;
- Proxy45 mask;
- Full Field FEM exposure;
- 1.35 NA, C-Quad 20 deg. illumination with Sigma 0.90/0.65, X-Y polarized.

We found that the ASML resist process has sufficient contrast to allow printing the 45nm dense lines and gives sufficient DOF to print isolated lines at 45nm too, see Figure 1

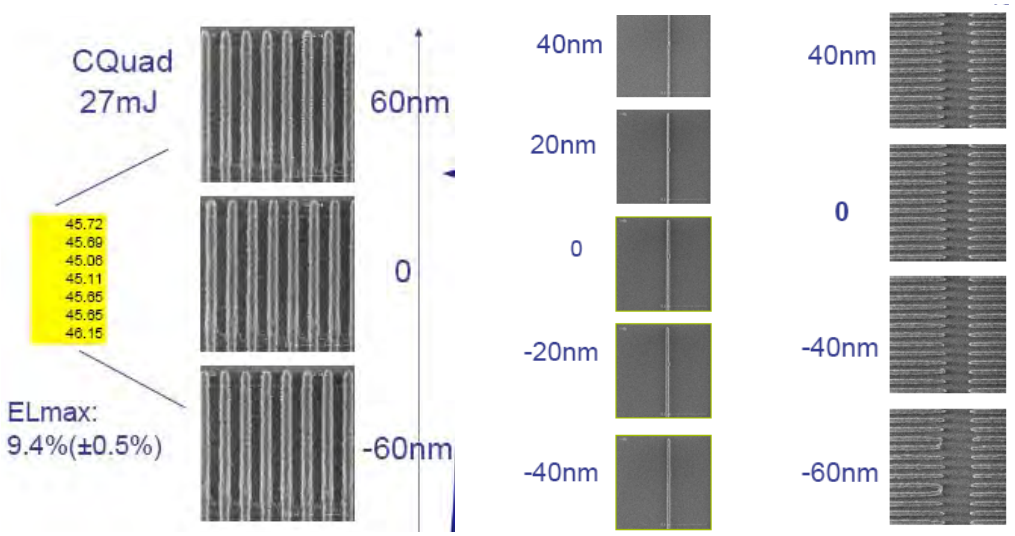


Figure 1 DoF and EL for the 45nm Dense and Iso lines and End of line pullbacks printed with the ASML resist process and C-Quad Illumination

Tool and process characterization for the OPC experiment

Each individual machine performs slightly different from other machine of the same type. Therefore machine data is measured which characterizes its performance.

We have collected the most part of the used TWINSCAN XT:1900i tool parameters to use them in the OPC modeling:

- E95 laser bandwidth;
- Pupil through slit;
- Aberrations through slit;
- Jones pupil;
- Vibrations in the XY and Z directions;
- Degree of polarization.

Ex. 1016, Page 387

Wafer and mask stacks were used for the calibration.

Feature Choice

Feature choice for OPC calibration depends on many inputs:

- Type of the device to calibrate (Flash, DRAM, Logic etc.);
- Design Rules of the device (min line width, space width, CD distribution on the design etc.);
- Process capabilities (forbidden pitch, biases etc.).

For the described OPC calibration we have chosen a Flash-type design so the target features are 1D dense lines although semi-Iso, Iso and 2D features were also considered.

The gauges that are used for the actual model regression and fitting have to meet the following requirements:

- Need to be design relevant: It does not make sense to have gauges that are beyond the design rule space. Gauge types should be in the present in the right ratio: If line through pitch model behavior is by far the most critical parameter, most gauges should be lines through pitch structures.
- Gauges need to be well measurable.
- Some gauges need to be sensitive towards parameters that are used in model regression, else singularities could arise, and the model may not have good predictive power on structures other then the calibration structures.

The initial selection of the features on the reticle contained pitches from 90nm to Iso and line widths from 45nm to 1000nm with and without AFs and End of Line (EOL) gaps with various pitches, line widths and gap sizes to cover design rule space. These structures were filtered to reduce CD measurement noise due to the low contrast for poorly printed patterns. The filters were:

- Wafer CDs of the 1D L/S with and without AFs: features with CDs below 35nm were excluded from the gauge.
- Printability: ILS was simulated for all structures by Aerial Image based Tachyon model and features with ILS below threshold were excluded from the gauge.

The final gauge set contained about 1200 1D structures and 300 2D structures to measure at the nominal Dose and Focus conditions. One hundred focus sensitive features were selected for the CD-SEM measurements through Process Window to develop the FEM OPC model.

CD-SEM Measurements and CD Data Analysis

CD data collection and analysis for the OPC experiment included:

- Automatic SEM recipe creation,
- CD-SEM measurements through the Process Window,
- Analysis and initial automatic filtering of the raw SEM data per dose and focus with various filters to capture poor measurements, misprinted features, SEM errors etc.

Ex. 1016, Page 388

- Automatic filtering through dose and focus.
- Final manual review of some problematic features basing on the automatic filtering results and actual OPC fit.

Statistics for the CD-SEM data collection

It sounds not very logical to achieve a model fit lower than the measured CD deviations, that's why advanced Lithography needs low noise CD-SEM measurements.

The number of repetitions for CD-SEM measurements was chosen according to observed CD deviations, required error range with 95% confidence and the formula below (see [1] for the reference):

$$ErrorRange = \overline{W} \pm \frac{1.96}{\sqrt{N}}\sigma \quad (1)$$

Where N is measurement times, σ is standard deviation of the measurements, W is average.

So if the required error range is below 0.5nm and the observed SE-SEM std. deviation is about 0.6-0.7nm, the required N should be more than 7. Unfortunately we expect some SEM measurements to be filtered due to noise, SEM errors, line roughness etc. so we have chosen N=10…15 for 1D features depending on the availability on the reticle and to avoid resist shrink due to multiple measurements on the same feature. The number of CD-SEM measurements for 2D features was less because of their availability on the mask. This is however not a critical issue due to the lower requirements on the 2D fit.

Automated CD-SEM recipe generation

As the number of individual features which were to be measured on the exposed wafers was very large (~1500), the choice was made to implement an automatic SEM recipe generation procedure. This was done to save time for the next cases: writing scripts and generating Reticle data also takes some time but the SEM recipe generation for the next calibration could be done in minutes. The automated SEM recipe creation also ensures that all SEM measurements on wafer were done in a systematic and predictable way, without the chance of wrongly measured gauges, due to operator errors when creating the recipes.

The SEM recipe generation works roughly in the following way:

- Import the Tachyon gauge file used for calibration;
- Import Reticle data file;
- Define in script whether of not differential movement by SEM is enabled;
- Make Recipe Build files used by Hitachi CD-SEM;
- Send files to CD-SEM with 'Recipe Builder' functionality.

CD-SEM measurements and analysis

CD-SEM measurements were done with the CG4000 Hitachi CD-SEM installed in the ASML Demo Lab in Veldhoven, The Netherlands. The measurements took about a day.

On some occasions measurement point image recognition fails on the CD-SEM. It then measures space iso lines or AF's instead of main features. When using well-made templates, this is a rare event. To filter wrong measurements and poorly printed

Ex. 1016, Page 389

features several tests were carried out based on the raw data, calculated averages and standard deviations:

- The WrongImageRec and DeltaSigma filter: When doing multiple measurements per feature, the one time image recognition fails, the given CD has a value clearly different than the other, rightly measured ones. Filtering these values has been implemented in the following way: First the standard deviation is calculated. Next, in turn every single (SEM FoV) data point for that gauge is removed and the standard deviation is recalculated. If the newly calculated sigma drops by a certain factor (for example 2), the corresponding measurement is considered wrong, and skipped for the calculation of a new average and sigma.
- The AbsoluteCD filter: If an individual data point (before averaging) has a value higher or smaller than certain reference values, the data point is removed altogether.
- The AF maxCD filter: For AF 1D structures only: If an individual data point (before averaging) has a value which is larger than a certain fraction (i.e. 0.45) of the half pitch, the data point is considered invalid and removed altogether.
- The Instability filter: For those points where more than x% of the data points are removed or less than a certain number of individual measurements is left, the feature is excluded from calibration.
- The finalSigma filter: This last filter puts a threshold on the final standard deviation.
- The DeltaSigma “cut-off” filter. This test first calculates the standard deviation from all individual data points left. Then it removes all data points which are more than N times the standard deviation from the mean value. Typically we use 1Sigma – 2Sigma filtering.

We also used the “FEM GaugeSet Check” tool from Brion Technologies. The application checks the gauges with the same feature (e.g. same CD and pitch) through focus and dose. In addition, the gauges can be grouped by the user so that these groups are also checked for outliers. For through-focus a polynomial fit is made on the wafer_CD versus delta_focus and the focus weight mask is obtained from several aspects, such as position of curve tip, shape of curve, fitting correlation etc.

The analysis script then adds the data to the Tachyon gauge file.

The weight of the data points was chosen according their type (1D critical DRs was higher than 1D not critical and higher than 2D), calculated std deviation of the CD-SEM measurements and simulated ILS of the features.

Model calibration results and discussion

Tachyon model calibration [2] was performed on the CD data and tool data listed above. Before the actual calibration the study on the RMS sensitivity to the tool and process parameters and calibration settings was performed on the available tool data and 1D CD data through PW.

The calibrations were performed in the following way:

- Parameters were set according to the reviewed case (see Table 1).
- Several resist model parameters were examined; the best one was used for the final resist modeling.
- The same ranges of parameters were used to start the calibrations; the parameters were tined till:
    - The best fit was achieved;
    - Reasonable physical boundaries of the fitting parameters were achieved;
    - The model became unstable.

The tool parameters could be preliminary ranked according their influence on the RMS of the reviewed TWINSCAN XT:1900i modeling. The **Error! Reference source not found.** below shows decreases of the modeling RMS per tool parameter, Aerial Image model (except the "complex resist model" one):

Table 1 Modeling parameters ranking

Parameters	Ranking: RMS(parameter off) - RMS(parameter on)			
	1D @ best FE	Rank @ best FE	1D thr. PW	Rank thr. PW
Complex resist model	6.8	**1**	7.1	**1**
3D mask	2.66	**2**	3.3	**2**
Parameterized Pupil	1.8	**3**	1.34	**3**
Jones Pupil	0.53	**4**	0.5	**4**
Vibrations	0.46	**4**	0.47	**4**
Meas. pupil	0.41	**4**	0.39	**4**
DoP	0.23	**5**	0.21	**5**
Laser E95	0.16	**5**	0.16	**5**

This ranking of the tool and process parameters could greatly help to optimize the data collection process, reduce the required tool time for tests, make the OPC model simpler without visible degradation of the model quality and also decrease the modeling runtime required for trials of different parameter combinations.

Unfortunately calibration with enabled measured (at the corresponding slit location) aberrations did not improve the fit, even made it a bit worse. We are studying this phenomenon and look forward to have an update soon.

As expected, the modeling RMS shows sensitivity to the std deviation of the input CD data. So, the RMS of the model calibrated with all available tool parameters, advanced resist model and 1D CD data filtered (see the DeltaSigma "cut-off" filter)

Ex. 1016, Page 391

with 1Sigma filter was smaller by 0.2nm compared to the similar calibration based on the 1D CD data filtered with 2Sigma.

Based on the RMS sensitivity study results we used “the efficient calibration practice” to achieve the best possible fit for existing conditions: all available tool and process parameters and most noise-free CD data filtering.

The results are:

- The RMS for the critical 1D L/S features (without AFs) at the best FE conditions is 0.94 nm;
- The RMS for the 1D features (with and without AFs) at the best FE conditions is 1.08nm;
- The RMS for the 1D features (with and without AFs) through PW is below 3nm, typically about 1.5-2nm;
- The RMS for the 2D features (EOL pullbacks) at the best FE conditions is about 3.5nm.
- The modeling coefficients are stable according the corresponding parameter limits.

1D features show very good fits, see Figure 2 (at pitch=0 the fit for completely isolated features is shown) and Figure 3. 2D features have a slightly worse fit, up to 3.5nm RMS but this is less critical because of the device nature for OPC (FLASH-like) and could be explained by less weights and more CD-SEM noise for 2D features in this experiment.

The model also shows good results for the through focus prediction, see Figure 4.

The simulated resist profiles of 2D features show excellent overlap with the corresponding SEM images but some improve for EOLs is still expected see Figure 5.

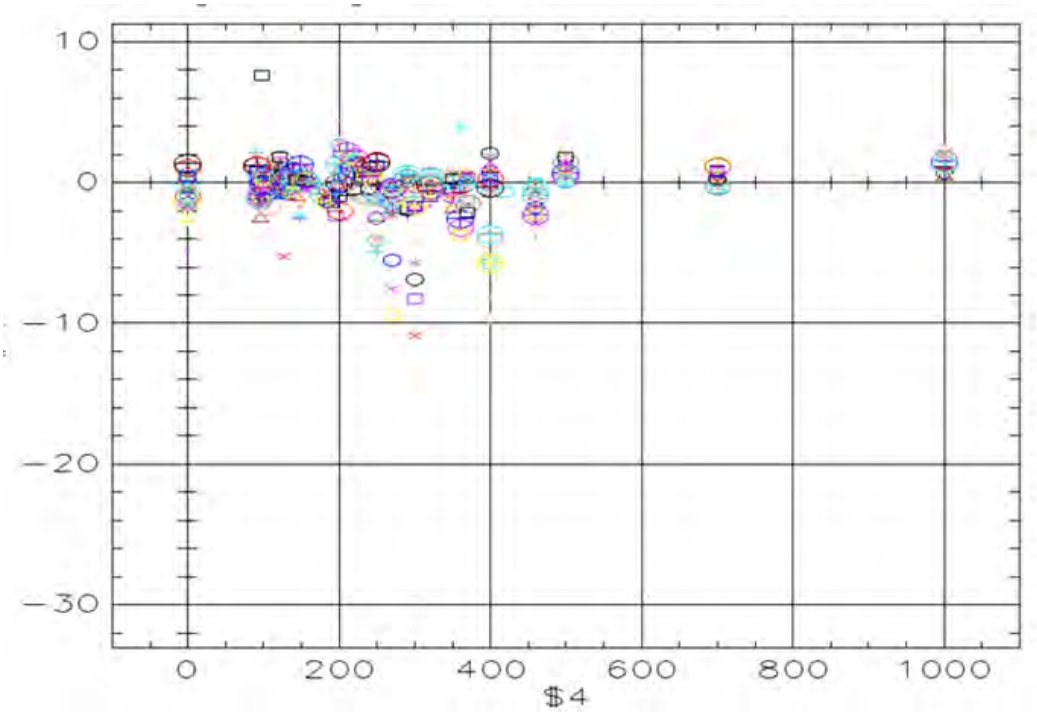


Figure 2 1D features RMS through pitch

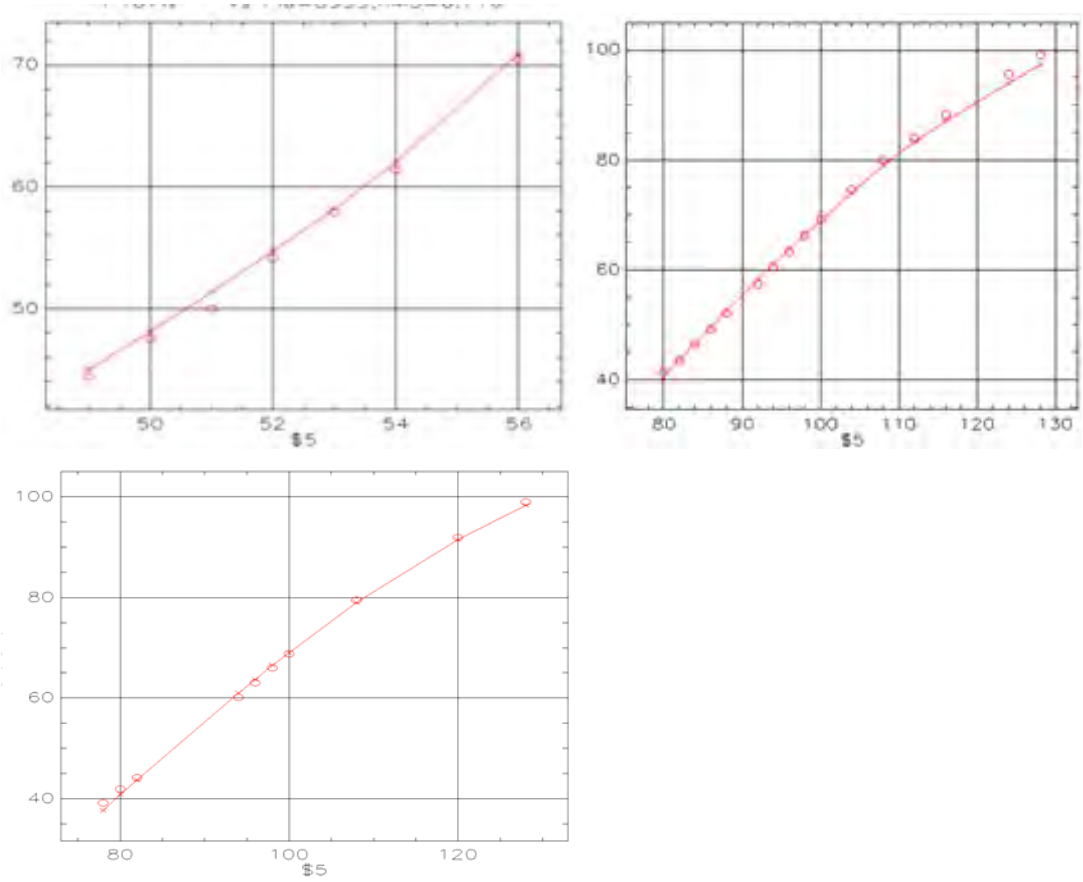


Figure 3 107nm, 350nm pitches and iso lines fit through LW with RMS 0.716, 0.839 and 0.751 respectively

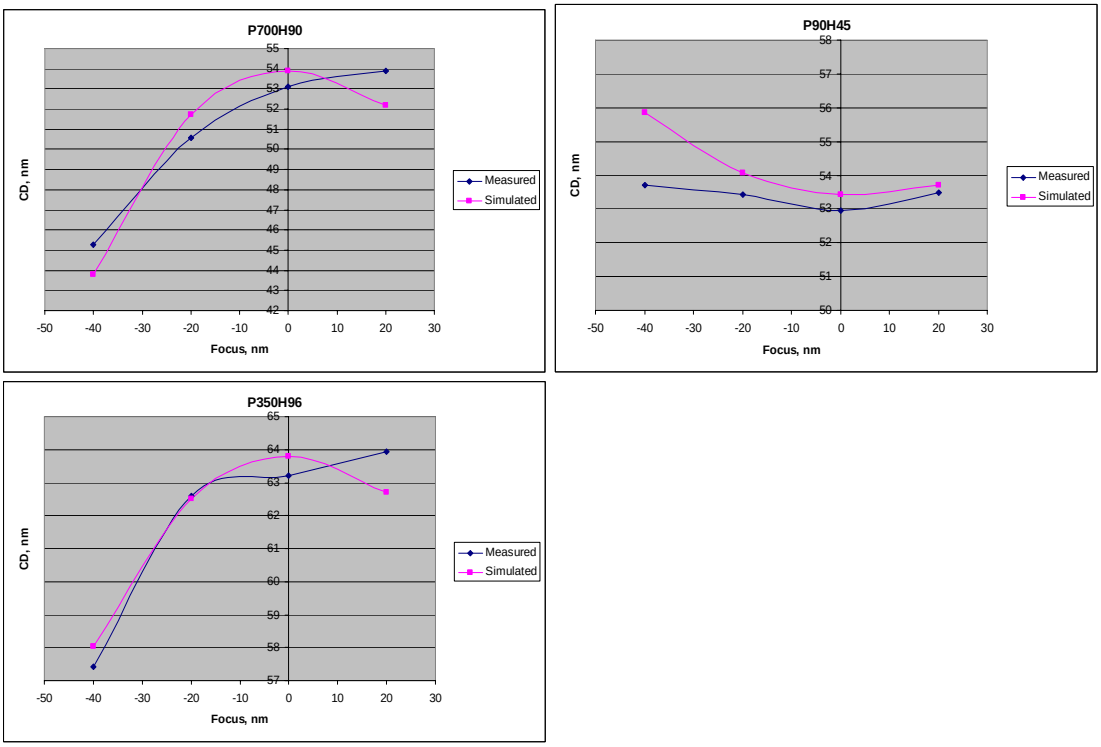


Figure 4 Through focus curves

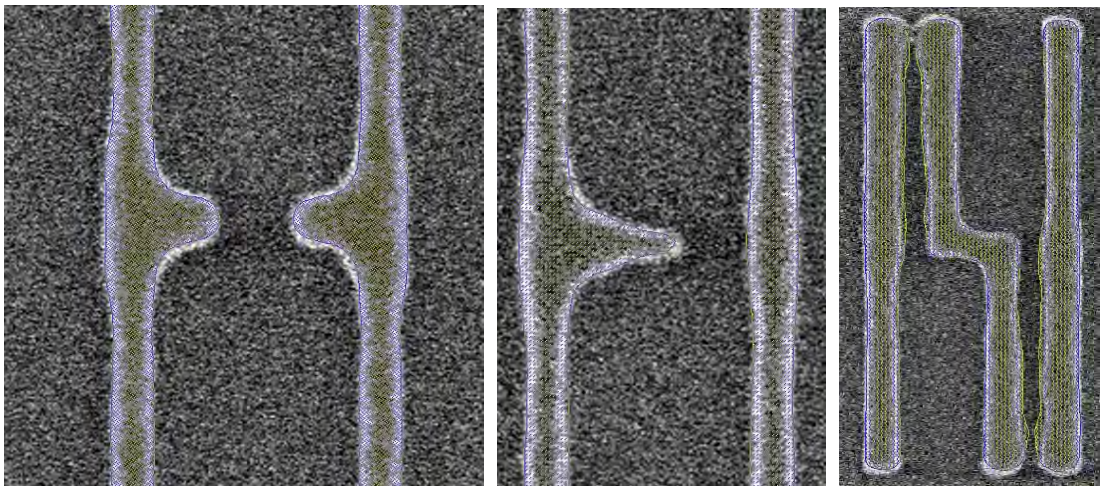


Figure 5 Simulated vs. CD-SEM profiles

Conclusion

We achieved high OPC modeling accuracy below 1nm for 1D features required for the advanced 45nm technology node and below.

The enabling tools for this achievement are the innovative ASML TWINSCAN XT:1900i scanner, Tachyon OPC software and developed data collection and preparation procedures.

The improved workflow and algorithms for the OPC experiment preparation, implementation and data analysis, tool and process parameters ranking for calibration allows to ASML:

- Early and fast OPC model development leading to the significant product-to-market time reduction;
- Reducing overall CDU and improving yield by achieving best OPC accuracy for advanced technologies at 45nm and below.

References

[1] Geng Han, Andrew Brendler, Scott Mansfield, and Jason Meiring, “Statistical Optimization of Sampling Plan and Its Relation to OPC Model Accuracy” Proc. SPIE Vol.6518 2007.

[2] Luoqi Chen; Yu Cao; Hua-yu Liu; Wenjin Shao; Mu Feng; Jun Ye, “Predictive focus exposure modeling (FEM) for full-chip lithography”, SPIE Vol. 6154, 2006.

ECS Transactions, 18 (1) 367-372 (2009)
10.1149/1.3096473 © The Electrochemical Society

Hot-Spots Aware Inverse Lithography Technology

YANG Yiwei, SHI Zheng, SHEN Shanhu, XIE Chunlei

Institute of VLSI Design, Zhejiang University, Hangzhou, 310027, China

This paper presents a new method of Inverse Lithography Technology (ILT) to accelerate its optimization process and convergence, which involves two steps in iteration: the first step is to find hot-spots in the layout in the optimization process and the second step is to optimize the hot-spots area. Information in optimization process such as gradient can be used in the first step to help finding the hot-spots. Optimized surroundings have been used as environment for local hot-spots fixing in the second step. In order to reduce ripple effects, specific penalty terms are used in the second step, thus the local fixing for hot-spots would minimally affect the surroundings through the procedure. This technique only focuses on the most un-optimized places thus accelerate the optimization process. The experiments show that this hot-spots aware ILT outperforms traditional ILT.

Introduction

As the 193nm wavelength applied to the technology nodes from 90nm to 32nm, pattern distortion in lithography becomes more and more severe. Using Resolution Enhancement Technologies (RET) such as Optical Proximate Correction (OPC), Assistant Feature (AF) and Phase Shift Mask (PSM) is no longer sufficient to satisfy the IC manufacturing requirements.

Inverse Lithography Technology (ILT) has the capability to push the lithography to its limit (1). It considers the mask optimization work as an inverse problem, and methods such as level set methods (1), gradient descent methods (2, 3) and simulated annealing algorithm can be used to solve it. Take the gradient related optimization methods as an example to show how the traditional ILT works. The cost function is a discrepancy measurement between the output contour and target contour. Universal format is as equation [1] shows, Litho() is lithography system function which contains develop process and etch process, and function distance() describes the discrepancy in contour space.

$$\text{cost\_function} = \text{distance}(\text{Litho}(\text{mask}), \text{target\_contour}) \quad [1]$$

Mask manufacturing factors and mask complex penalty terms have been posed on the mask space. To find the optimal mask that makes the cost_function minimum, the mask space needs to be searched iteratively. The iteration is moving an optimal distance from one starting point $mask_k$ along the direction opposite to gradient-related direction to a new point $mask_{k+1}$. The procedure is described by equation [2], where grad() means the gradient-related direction, and s is the distance.

$$mask_{k+1} = mask_k - grad(\text{cost\_function})*s \quad [2]$$

In some cases at the optimization process (shown in Figure 1) the cost function curve becomes flatten and converges slowly (4) and most areas of the mask have very small cost function value except dispersed hot-spots. Generally speaking, hot-spots represent the places which do not pass the Design for Manufacturing (DFM) verification, and here

Ex. 1016, Page 395

hot-spots symbolize the places with large cost function value or gradient value, which have large possibility not passing DFM verification. To fix these hot-spots, the whole mask is re-calculated during the backend of the optimization process, which costs much time; and even worse the cost function may stick at a local minimum thus the hot-spots can not be fixed. This phenomenon not only delays the progress of optimization but also severely influences the convergence, and in some cases it leads to un-qualified results. This paper proposes a new method to accelerate the optimization process and convergence, which can be used in other optimization methods (such as simulated annealing) analogously. It selects the hot-spots and re-optimizes them locally with penalty terms to restrain ripple effects, which also can make cost function jump out of the local minimum to achieve a better minimum due to a relatively small mask space. Optimization theory indicates that if a global minimum is achieved, the cost function must be minimum in the sub-space, which means that fixing hot-spots can accelerate convergence.

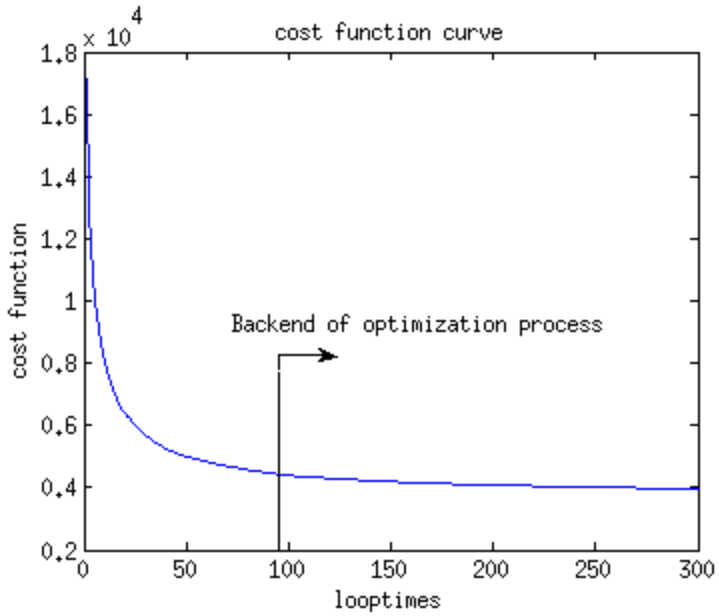


Figure 1. Curve of cost function in optimization process.

In the second section the new method that accelerates optimization process is proposed and the third section is for experiments and discussion.

Hot-Spots Aware ILT

As stated above, further computation spend in the optimized parts of the mask is not necessary. Optimize the dispersed hot-spots only and at the same time make the result correct and convergent is a way to accelerate the optimization process. To accomplish this, two challenges need to be solved first. The first challenge is how to locate the hot-spots and the second is how to reduce the ripple effects, which is induced by fixed hot-spots.

To locate the hot-spots, the information such as cost function or gradient in optimization process is sufficient. Optimization theory indicates that in an epsilon neighborhood of the minimum point the gradient is smaller than a given value. So if a

Ex. 1016, Page 396

point has a large cost function value or large gradient value, it is far to be optimized. The weighted sum of gradient and cost function can be the judging function for hot-spots, as equation [3] shows, in which t_1 and t_2 are the weighting factors.

$$judging\_function = t_1*gradient+t_2*cost\_function \quad [3]$$

Choosing the optimized surroundings of the hot-spots as the environment, and only hot-spots and its environment participate in the optimization, thus reduce the computation. In practice, the hot-spots are not points but small areas, and the judging function is processed on that hot-spots area correspondingly. The judgement of hot-spots needs other constraints in order to filter out the fake hot-spots and leave bridge and break hot-spots only. Figure 2 shows the distribution of judging function over a mask area. The areas under closed to many peaks are hot-spots.

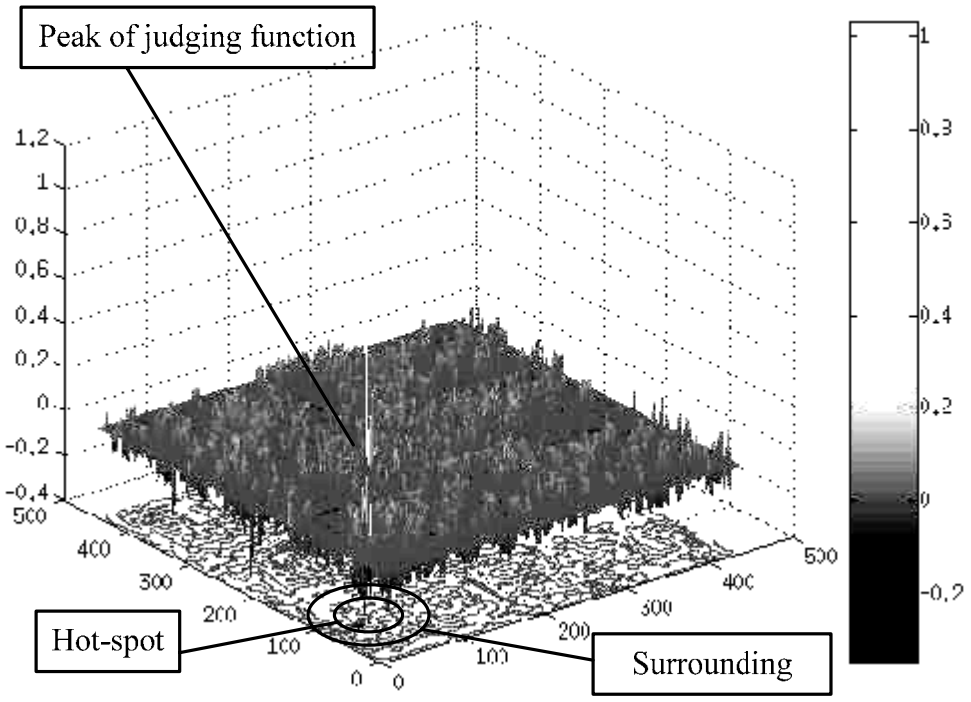


Figure 2. Distributions of the judging function over a mask area and the corresponding hot-spots. The area under the peak is hot-spot.

The regions which do not participate the optimization will be affected by the fixed hot-spots. This so-called ripple effects has limited direct impact range due to the limited ambit of lithography kernel, but the indirect impact range can be long (5). Ripple effects may produce disastrous influence on the optimization result, and it stops propagating when a change of a polygon does not change its neighboring polygons. Based on this fact, to reduce ripple effects, carefully designed penalty terms are added in local hot-spots ILT. The role of the penalty terms is to guarantee that the local optimization direction is coincident with the global optimization direction, which means the environment which participates in local ILT converges to its former self. Here, former self means the surroundings chosen as the environment participated in local ILT, as Figure 3 shows. ripple effects are strong at the inner of the surroundings but weak out of the surroundings due to rarely changed surroundings. The surroundings are walls to block ripple effects and it helps to renew the area of hot-spots without errors. The width of surroundings needs carefully choosing and if necessary, multi-step-surroundings can be used.

Ex. 1016, Page 397

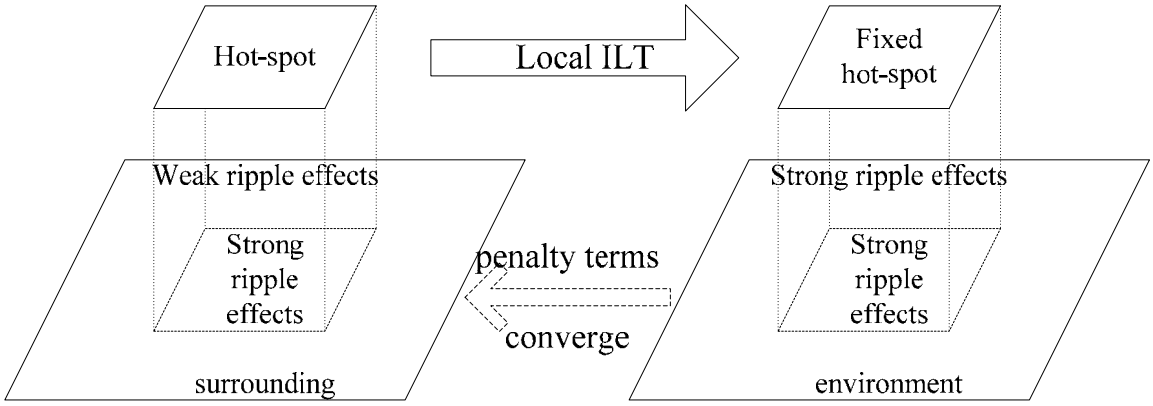


Figure 3. Schematic diagram of local hot-spot ILT. Hot-spot is fixed after local ILT and the environment convergence to its former self after local ILT to reduce ripple effects.

One thing worthy of note is that when to start hot-spots aware ILT is not fixed. Usually the slop of the cost function and sum of hot-spots areas are good foundation to determine the best time.

Experiments and Discussion

Experiments are designed to test the speed and quality of the hot-spots aware ILT. The parameters of the experiments, including kernel parameters and optimization parameters, are only designed for experiments. Kernels are jinc functions, and lambda = 193nm, traditional illumination with s = 0.8, NA = 0.75, threshold = 0.27, grid = 7nm. Five different layouts of 90nm are used in experiments.

TABLE I. Run time of traditional ILT and hot-spot aware ILT.

Layout number	Traditional ILT (second)	Hot-spots aware ILT (second)	Speedup ratio
Layout 1: memory	397	359	1.11
Layout 2: contact	137	117	1.17
Layout 3: logic1	498	473	1.05
Layout 4: logic2	366	352	1.03
Layout 5: logic3	713	659	1.08

Table I. shows that hot-spots aware ILT spend less time than traditional ILT by about 10%. The biggest speedup ratio is 1.17, and it is not only connected with the optimization parameters and layout style, but also related with the hot-spots number. The saved time is from the final 40% ~ 50% optimization process, and if the hot-spots are sparsely distributed in the mask the speedup ratio is satisfactory. To achieve the max speedup ratio, the optimization parameters need carefully choosing.

TABLE II. EPE of traditional ILT and hot-spot aware ILT

Layout number	Traditional ILT (grid)	Hot-spots aware ILT (grid)
Layout 1: memory	1.2	1.1
Layout 2: contact	0.5	0.6
Layout 3: logic1	1.2	1.2
Layout 4: logic2	0.9	1.0
Layout 5: logic3	0.8	0.8

Ex. 1016, Page 398

Table II shows the quality of the results of the two different ILTs. The grid is 7nm, thus the max EPE is about 8.4nm, which is acceptable. No significant difference of the two kinds of results can be seen in this Table II. Ripple effects are efficiently restrained.

TABLE III. Number of hot-spots of traditional ILT and hot-spot aware ILT

Layout number	Traditional ILT	Hot-spots aware ILT
Layout 1: memory	0	0
Layout 2: contact	0	0
Layout 3: logic1	1	0
Layout 4: logic2	0	0
Layout 5: logic3	3	1

Table III is the number of hot-spots of the two different ILTs. The results show that the hot-spots, which stick at a local minimum, can be fixed in local ILT due to its smaller solution space. Layout 5 indicates that not all hot-spots can be fixed in local ILT, and using better optimization methods might be a good alternative.

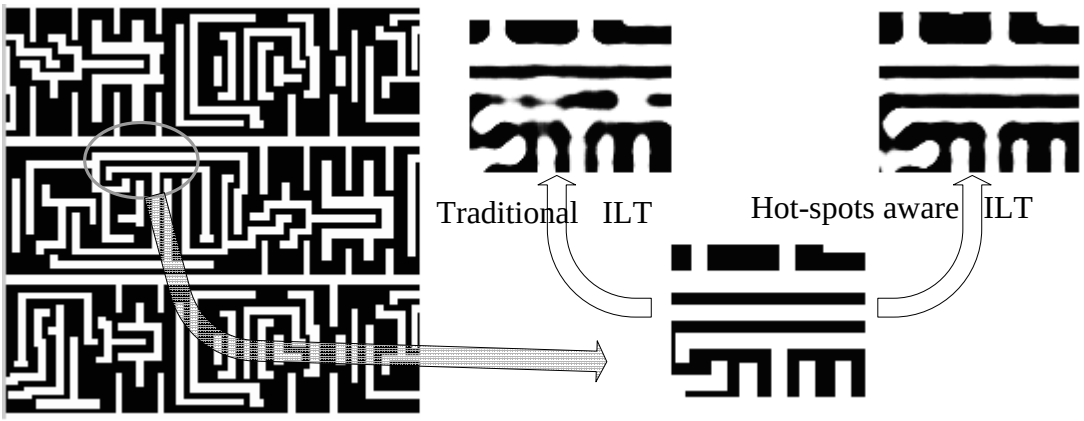


Figure 4. Hot-spots aware ILT eliminates hot spot

Figure 4 shows that the optimization sticks at a local minimum with a hot-spot. After hot-spots aware ILT, the hot-spot is fixed.

In the mean while, during hot-spots aware ILT process, the most error-prone places are reported after optimization process, which helps to verify the mask.

Conclusions

This paper proposes a hot-spots aware ILT, which accelerates the optimization process and convergence, and meanwhile restrains the ripple effects. Experiments results show that hot-spots aware ILT outperforms traditional ILT. Most important, the hot-spots aware ILT can make the cost function, which may stick at a local minimum, jump out to a better minimum.

References

1. Linyong Pang, Yong Liu and Dan Abrams, in International Semiconductor Technology Conference/2007, Ming Yang, Tomi Li, Fei Xu, Qingyuan Han and G.S. Mathad, Editors, PV 2007-01, p. 650, Proceedings of the 6th International Conference on Semiconductor Technology (2007).

Ex. 1016, Page 399

2. Yuri Granik, Journal of Microlithography, Microfabrication, and Microsystems, Vol. 5, 043002 (Dec. 13, 2006).
3. Amyn Poonawala and Peyman Milanfar, IEEE Transactions on Image Processing, Vol. 16, No. 3, March 2007.
4. Shanhu Shen, Peng Yu and David Z. Pan, in Photomask Technology 2008, Hiroichi Kawahira Larry S. Zurbrick, Editors. Vol. 7122, 712241, Proc. of SPIE.
5. Hongbo Zhang and Zheng Shi, in International Semiconductor Technology Conference/2007, Ming Yang, Tomi Li, Fei Xu, Qingyuan Han and G.S. Mathad, Editors, PV 2007-01, p. 781, Proceedings of the 6th International Conference on Semiconductor Technology (2007).

ECS Transactions, 18 (1) 373-379 (2009)
10.1149/1.3096474 © The Electrochemical Society

Trefoil Aberration Measurement of Lithographic Projection Optics Based on Linewidth Asymmetry of the Aerial Image

Qiongyan Yuan [a,b], Xiangzhao Wang [a,b], and Zicheng Qiu [a,b]

[a] Information Optics Laboratory, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China
[b] Graduate School of the Chinese Academy of Sciences, Beijing 100039, China

The degradation of critical dimension uniformity caused by trefoil aberration of lithographic projection optics is a serious problem in production of dynamic random access memory devices. In the present paper, we propose a novel method for measuring the trefoil aberration of lithographic projection optics. The trefoil aberration can be extracted from the linewidth asymmetry of the aerial image of brick wall patterns. The present method has an advantage of timesaving. The test mark used in the present method is easy and cheap to fabricate, and the model for aberration extraction is simple. The PROLITH simulation results show that, the trefoil aberration can be measured with high accuracy.

Introduction

The projection optics system is one of the most important systems in a step-and-scan lithographic tool. Wavefront aberrations of the projections optics cause a degradation of the image quality as well as a significant reduction of process window of the lithographic process(1-3). Recently, the degradation of critical dimension (CD) uniformity across the exposure field caused by lens aberrations of a lithographic tool has become one of the critical issues in production of semiconductor devices, especially in production of the dynamic random access memory (DRAM) devices[4]. The DRAM isolation pattern defines the active area of two DRAM transistors. The control of the size uniformity with the island area is crucial for the reliable operation of the device (5). The asymmetry of the DRAM isolation pattern is mainly influenced by trefoil aberrations which are represented by Zernike coefficients Z_{10}, Z_{11}, Z_{19} and Z_{20} (6). The impact of trefoil aberrations on printed patterns depends on pattern shape, pattern dimension, pattern orientation, illumination settings etc. As the design rule shrinks, degradation of CD uniformity across the exposure field caused by trefoil aberrations of the lithographic projection lens becomes more serious in the lithographic process. To predict and minimize the adverse effects on printed patterns, the semiconductor industry calls for fast and accurate in-situ measurement techniques for measuring the trefoil aberration.

In recent years, a number of resist-based methods for measuring the trefoil aberration of lithographic projection optics have been reported. Nomura et al. proposed a measurement method based on three-beam interference and pupil sampling(7-8). Trefoil aberration can be calculated by relative shifts of three marks oriented at 0°, 120° and 240°. The measurement marks are fabricated by double exposure, so the measurement procedure is complicated. Kirk et al. reported a measurement method based on two-beam interference (9). Trefoil aberration can be extracted from focus shifts of the aberrated images. However, the measurement accuracy of the aberrations is restricted because the

measurement accuracy of the focus shifts is low. Mingying Ma et al. also proposed a measurement method based on two-beam interference (10). The measurement accuracy of trefoil aberration was improved. However, the box-in-bars mark used in the method is difficult and expensive to fabricate. In addition, the methods mentioned above are not straightforward measurement techniques. They involve exposure of resist and rely on the formation of intermediate images in resist which are subsequently analyzed by a scanning electron microscope, an overlay inspection tool or an optical microscope. So these methods are time consuming.

In this paper, we propose a novel method for measuring the trefoil aberration of lithographic projection optics based on linewidth asymmetry of the aerial image. In this method, horizontal and vertical brick wall patterns are used as measurement marks. Trefoil aberration can be extracted from linewidth asymmetry of the aerial image of the measurement marks at multiple illumination settings. The linewidth asymmetry can be directly measured by an aerial image sensor which is built into the wafer stage. So the present method is a straightforward measurement method and has the advantage of timesaving. The test mark used in the present method is easy and cheap to fabricate, and the model for aberration extraction is simple. Using lithographic simulator PROLITH, the sensitivities of Zernike coefficients corresponding to trefoil aberration are calculated, and the measurement accuracy of trefoil aberration is analyzed.

Measuring Principle

Aberrations can be described as wavefront errors in the pupil plane of a lithographic projection lens. The shape of an ideal wavefront with no aberration is exactly spherical. Aberrations cause a deviation of the wavefront with respect to the ideal wavefront, leading to a shift and deformation of the image. The Zernike polynomials which represent the wavefront aberrations in the projection lens can be expressed as

$$\begin{aligned} W(\rho,\theta) &= \sum_{n=1}^{\infty} Z_n \cdot R_n(\rho,\theta), \qquad n \in Z \\ &= Z_1 + Z_2\rho\cos\theta + Z_3\rho\sin\theta + Z_4(2\rho^2-1) + \cdots + Z_{10}\rho^3\cos 3\theta + Z_{11}\rho^3\sin 3\theta \quad (1) \\ &\quad + \cdots + Z_{19}(5\rho^2-4)\rho^3\cos 3\theta + Z_{20}(5\rho^2-4)\rho^3\sin 3\theta + \cdots \quad , \end{aligned}$$

where ρ is the normalized radius of the exit pupil, and θ is the azimuthal angle. Z_{10} represents trefoil rotated 30°. Z_{11} represents trefoil. Z_{19} represents high-order trefoil rotated 30°. Z_{20} represents high-order trefoil.

The measurement mark used to measure trefoil aberration in the present method is shown in Fig. 1. The horizontal brick wall pattern used to measure Z_{10} and Z_{19} is shown in Fig. 1(a), and the vertical brick wall pattern used to measure Z_{11} and Z_{20} is shown in Fig. 1(b). The horizontal and vertical brick wall patterns have the same structures and dimensions, but different orientations. The feature width, feature height, X-pitch, Y-pitch of the horizontal brick wall patterns are 600nm, 130nm, 1000nm, 520nm, respectively, after an optimization of sensitivities to trefoil aberration. Trefoil aberration affects the intensity distribution of the aerial image of the measurement mark, and leads to CD asymmetry. The CD asymmetry can be defined as

Ex. 1016, Page 402

$$A_y = L_1 - L_2, \tag{2}$$

$$A_x = L_3 - L_4, \tag{3}$$

where L_1, L_2, L_3 and L_4 are the linewidth of the aerial image at given measurement threshold. The structure of the diffraction spectrum of the measurement is honeycomb pattern. The diffraction spectrum has its orders positioned in the regions where large phase errors are introduced by trefoil aberration, and consequently the sensitivity to CD asymmetry is large.

The CD asymmetry caused by trefoil aberration depends on pattern shape, pattern dimension, pattern orientation, illumination settings etc, because the intensity distribution of diffraction light varies with these parameters. The amplitude of the aberration-induced effects can be modified by imaging measurement marks with different shapes, dimensions, orientations, and by changing the numerical aperture (NA) and partial coherence (σ). Using the last two variables is the most convenient, since modern scanners have the flexibility to automatically set NA and σ. We do not need to fabricate measurement marks with different structures and dimensions. So the cost of the measurement is saved.

Z_{10} and Z_{19} affect the CD asymmetry A_y. Z_{11} and Z_{20} affect the CD asymmetry A_x. The CD asymmetry is approximately linear with the Zernike coefficients corresponding to trefoil aberration. For horizontal structure shown in Fig. 1(a), CD asymmetry caused by Z_{10} and Z_{19} can be expressed as

$$A_y(NA,\sigma) = S_1(NA,\sigma) Z_{10} + S_2(NA,\sigma) Z_{19}, \tag{4}$$

and for vertical structure shown in Fig. 1(b), CD asymmetry caused by Z_{11} and Z_{20} can be expressed as

$$A_x(NA,\sigma) = S_3(NA,\sigma) Z_{11} + S_4(NA,\sigma) Z_{20}, \tag{5}$$

where $A_y(NA,\sigma)$ and $A_x(NA,\sigma)$ are the CD asymmetry at the given NA and partial coherence setting, which can be measured by an aerial image sensor, and $S_1(NA,\sigma)$, $S_2(NA,\sigma)$, $S_3(NA,\sigma)$ and $S_4(NA,\sigma)$ are the sensitivities and can be expressed as

$$S_1(NA,\sigma) = \frac{\partial A_y(NA,\sigma)}{\partial Z_{10}}, \tag{6}$$

$$S_2(NA,\sigma) = \frac{\partial A_y(NA,\sigma)}{\partial Z_{19}}, \tag{7}$$

$$S_3(NA,\sigma) = \frac{\partial A_x(NA,\sigma)}{\partial Z_{11}}, \tag{8}$$

$$S_4(NA,\sigma) = \frac{\partial A_x(NA,\sigma)}{\partial Z_{20}}. \tag{9}$$

The sensitivities vary with the NA and partial coherence settings, and can be calculated by lithographic simulator such as PROLITH, Solid-C, etc.

For each NA and partial coherence setting, two linear equations are obtained. With multiple NA and partial coherence settings, two sets of equations can be written as

$$\begin{bmatrix} A_y(NA_1,\sigma_1) \\ A_y(NA_2,\sigma_2) \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} \frac{\partial A_y(NA_1,\sigma_1)}{\partial Z_{10}} & \frac{\partial A_y(NA_1,\sigma_1)}{\partial Z_{19}} \\ \frac{\partial A_y(NA_2,\sigma_2)}{\partial Z_{10}} & \frac{\partial A_y(NA_2,\sigma_2)}{\partial Z_{19}} \\ \vdots & \vdots \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} Z_{10} \\ Z_{19} \end{bmatrix}, \tag{10}$$

$$\begin{bmatrix} A_x(NA_1,\sigma_1) \\ A_x(NA_2,\sigma_2) \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} \frac{\partial A_x(NA_1,\sigma_1)}{\partial Z_{11}} & \frac{\partial A_x(NA_1,\sigma_1)}{\partial Z_{20}} \\ \frac{\partial A_x(NA_2,\sigma_2)}{\partial Z_{11}} & \frac{\partial A_x(NA_2,\sigma_2)}{\partial Z_{20}} \\ \vdots & \vdots \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} Z_{11} \\ Z_{20} \end{bmatrix}. \tag{11}$$

Equations (10) and (11) are over-determined and can be solved by the least square method. Using the aerial image sensor, the CD asymmetry of aerial image of the measurement mark in different field positions are measured at multiple illumination settings. Zernike coefficients Z_{10}, Z_{11}, Z_{19} and Z_{20} in different field positions can be determined.

PROLITH Simulation

PROLITH is a well known lithographic simulator and widely used in the lithographic industry as a research and development tool, to help provide a fundamental understanding of all aspects of the lithography process, and to validate and improve the industry's theoretical understanding of lithography, and to provide a tool to the average lithographer to apply this theory to real lithography problems (11-12).

Variation ranges of sensitivities are the key factors which influence the measurement accuracy of trefoil aberration. Using the lithographic simulator PROLITH, the variation ranges of sensitivities of Zernike coefficients corresponding to trefoil aberration were calculated at multiple NA and partial coherence settings, which are commonly used in mainstream lithographic tools.

The wavelength used in the simulation is 193nm. The variation range of the NA of projection optics is 0.65~0.85. The variation ranges of the partial coherence under conventional illumination and annular illumination are 0.3~0.8 and 0.3~0.7, respectively. The annular width is 0.3.

Ex. 1016, Page 404

Figures 2(a) and 2(b) show the sensitivities of Z_{10}/Z_{11} for the present technique at multiple NA and partial coherence settings, under conventional illumination and annular illumination, respectively. For all the illumination settings, the sensitivities of Z_{10}/Z_{11} vary from -2.63 to 1.66. The variation range of sensitivities of Z_{10}/Z_{11} is 4.29.

Figures 3(a) and 3(b) show the sensitivities of Z_{19}/Z_{20} for the present technique at multiple NA and partial coherence settings, under conventional illumination and annular illumination, respectively. For all the illumination settings, the sensitivities of Z_{19}/Z_{20} vary from –2.24 to 3.97. The variation range of sensitivities of Z_{19}/Z_{20} is 6.21.

In the present paper, the CD asymmetry of the aerial image of the measurement mark was used to extract the trefoil aberration. The measurement accuracy of the trefoil aberration depends on the measurement accuracy of the CD asymmetry. The measurement accuracy of trefoil aberration $MA_{trefoil}$ can be estimated as

$$MA_{trefoil} \propto \frac{MA_{CDA}}{\left|S_{\max} - S_{\min}\right|}, \tag{12}$$

where MA_{CDA} is the measurement accuracy of CD asymmetry, S_{max} and S_{min} are the maximum and minimum values of sensitivity.

From the simulation results of sensitivities, the measurement accuracy of trefoil aberration cab be estimated by Eq. (12), and the MA_{CDA} is 2nm. The simulation results of the sensitivities and the estimated measurement accuracy are summarized in Table 1. The estimated measurement accuracy of low-order trefoil aberration is 0.466nm. The estimated measurement accuracy of high-order trefoil aberration is 0.322nm. The measurement accuracy of trefoil aberration obviously increases, compared to the measurement method proposed by Kirk et al. and Mingying Ma et al. (10).

Conclusions

A novel method for measuring the trefoil aberration of lithographic projection optics based on linewidth asymmetry of the aerial image has been proposed. Horizontal and vertical brick wall patterns were used as measurement mark in the present method. The trefoil aberration can be extracted from the CD asymmetry of the aerial image of the brick wall patterns at multiple illumination settings, using the least square method. The present method has an advantage of timesaving. The test mark used in the present method is easy and cheap to fabricate, and the model for aberration extraction is simple. Using the lithographic simulator PROLITH, we have calculated the sensitivities corresponding to trefoil aberration at a series of illumination settings and analyzed the measurement accuracy. According to the simulation results, the measurement accuracy of trefoil aberration has obviously increased, compared with previous methods.

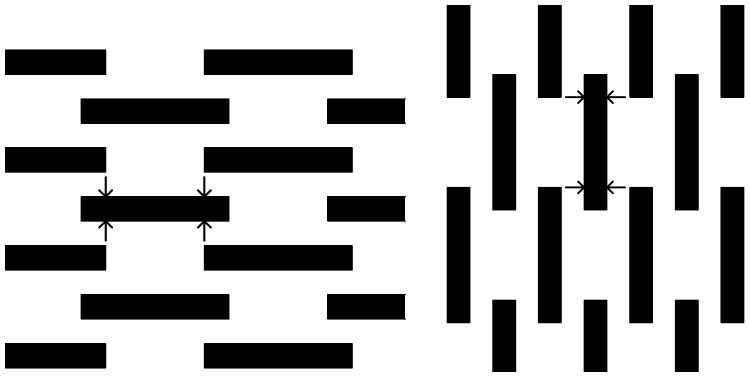
Ex. 1016, Page 405

Acknowledgments

This work was supported by a grant from the Key Basic Research Program of Science and Technology Commission of Shanghai Municipality NO. 07JC14056 and National Natural Science Foundation of China under Grant NO. 60578051. The authors would like to thank NERCLE (National Engineering Research Center for Lithographic Equipment, China) for the support of this work.

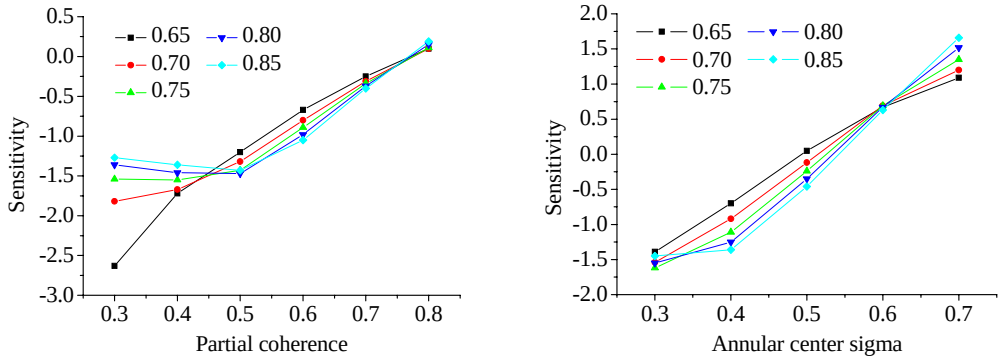
References

1. Q. Yuan, X. Wang, Z. Qiu, F. Wang, M. Ma, and L. He, "Coma measurement of projection optics in lithographic tools based on relative image displacements at multiple illumination settings," Opt. Express **15**, 15878-15885 (2007).
2. F. Wang, X. Wang, M. Ma, D. Zhang, W. Shi and J. Hu, "Aberration measurement of projection optics in lithographic tools by use of an alternating phase-shifting mask," Appl. Opt. **45**, 281-287 (2006).
3. P. Graupner, R. Garreis, A. Goehnermeier, T. Heil, M. Lowisch and D. Flagello, "Impact of wavefront errors on low k1 processes at extreme high NA," Proc. SPIE **5040**, 119-130 (2003).
4. D. Flagello, H. van der Laan, J. Van Schoot, I. Bouchoms, B. Geh, "Understanding Systematic and Random CD variations using Predictive Modelling Techniques," Proc. SPIE **3679**, 162-175 (1999).
5. V. L. Rideout, "One-Device Cells for Dynamic Random-Access Memories: A Tutorial," IEEE Trans. Electron Devices, **ED-26**, 839-852 (1979).
6. K. Kobayashi, T. Yao, Y. Yanagishita, I. Hanyu, "Impact of lens aberration on pattern symmetry of DRAM cells" Proc, SPIE **4000**, 1269-1280 (2000).
7. H. Nomura and T. Sato, "Techniques for measuring aberrations in lenses used in photolithography with printed patterns," Appl. Opt. **38**, 2800-2807 (1999).
8. H. Nomura, K. Tawarayama and T. Kohno, "Aberration measurement from specific photolithographic images: a different approach," Appl. Opt. **39**, 1136-1147 (2000).
9. J. P. Kirk, G. Kunkel, A. K. Wong, "Aberration measurement using in situ two-beam interferometry," Proc. SPIE **4346**, 8-14 (2001).
10. M. Ma, X. Wang and F. Wang, "Aberration measurement of projection optics in lithographic tools based on two-beam interference theory," Appl. Opt. **45**, 8200-8208 (2006).
11. C. Mack, "Lithography Simulation in Semiconductor Manufacturing," Proc. SPIE **5645**, 63-83 (2005).
12. C. Mack, "30 Years of Lithography Simulation," Proc. SPIE **5754**, 1-12 (2005).



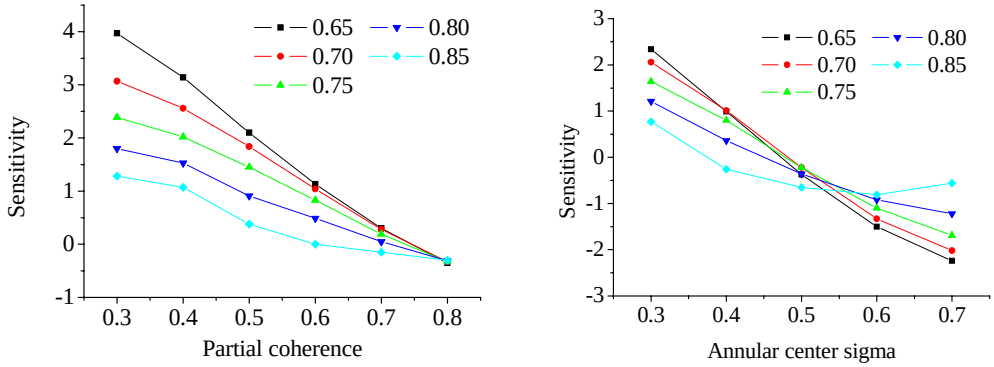
(a) (b)

Fig. 1. Schematic of the brick wall patterns, (a) horizontal, (b) vertical.



(a) (b)

Fig. 2. Sensitivities of Z_{10}/Z_{11} versus NA and partial coherence. (a) Conventional illumination. (b) Annular illumination



(a) (b)

Fig. 3. Sensitivities of Z_{19}/Z_{20} versus NA and partial coherence. (a) Conventional illumination. (b) Annular illumination

Table 1. Simulation Results of the Sensitivities and the measurement accuracy

Zernike	Maximum	Minimum	Variation range	Measurement accuracy (nm)
Z_{10}/ Z_{11}	1.66	-2.63	4.29	0.466
Z_{19}/ Z_{20}	3.97	-2.24	6.21	0.322

ECS Transactions, 18 (1) 381-390 (2009)
10.1149/1.3096475 © The Electrochemical Society

High-Accuracy Measurement of Coma with the Optimized Alternating Phase-Shifting Mask with a Specific Phase Width

Zicheng Qiu[a,b], Xiangzhao Wang[a,b*], Qiongyan Yuan[a,b], Fan Wang[c]

[a]Information Optics Laboratory, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China
[b]Graduate School of the Chinese Academy of Sciences, Beijing 100039, China
[c]National Engineering Research Center for Lithographic Equipment, Shanghai 201203, China
[*]Corresponding author. Address: Information Optics Laboratory, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, P.O. Box 800-211, JiaDing, Shanghai 201800, China. Tel.: +86 2169918692; fax: +86 2169918692. E-mail address: wxz26267@siom.ac.cn

The correlation between the coma sensitivity of the Alternating Phase-Shifting Mask (Alt-PSM) and the Alt-PSM's structure is studied. It is found that an optimized Alt-PSM whose phase width is two thirds of its pitch has a higher sensitivity to coma than Alt-PSMs with the same pitch and the different phase widths, when the pitch of the Alt-PSM makes only ±1 and ±3 orders diffraction light enter the lens pupil. The optimized Alt-PSM is used as a measurement mark to retrieve coma aberration from the projection optics in lithographic tools. In comparison with an ordinary Alt-PSM mark whose phase width is one half of its pitch, the measurement accuracies of Z_7 and Z_{14} have increased by 15% and 31% respectively.

Introduction

The projection lens system is one of the most important systems in a lithography tool. The coma aberration of the projection lens can lead to two kinds of image quality degradations. First, it can cause Image Placement Error (IPE), which leads to overlay errors. Second, it can cause line-width asymmetry, which degrades the Critical Dimension (CD) uniformity (1-6). In order to correct coma effectively, it is necessary to establish a fast and accurate in situ measurement technique for retrieving coma of projection optics. As the CD decreases, especially with the use of resolution enhancement techniques, coma measurement techniques with higher accuracy are needed. In the low-k_1 lithography process, for the high NA projection optics, its residual wavefront aberrations, including coma, have decreased to 10 mλ or less, and it is necessary to measure wavefront aberration on the exposure tool with an accuracy of 2mλ (5).

TAMIS (TIS at Multiple Illumination Settings) technique has been developed to measure wavefront aberrations (6). To determine the wavefront aberration, the aerial image of a binary mark is directly measured in this technique. Its measurement accuracy of coma reaches 2nm for ArF lithography tools. The measurement accuracy is determined by the measurement mark's sensitivity to coma. We have used an Alt-PSM as the measurement

Ex. 1016, Page 409

mark instead of the binary mark to improve the measurement mark's sensitivity to aberrations and the accuracy of the TAMIS method (1). However, the correlation between the sensitivity to coma and the structure of the measurement mark has been ignored by us. The ordinary Alt-PSM whose phase width (*pw*) is one half of its pitch (*p*) has been used as the measurement mark. The pitch and phase width of the mark have not been optimized to attain the highest sensitivity. In this paper, using the lithographic simulator PROLITH, we study the influence of the Alt-PSM's structure on its sensitivity to coma, and optimize the structure to get the highest sensitivity. Then the simulation results are discussed analytically. At last, the optimized Alt-PSM is used as a measurement mark to improve the accuracy of coma measurement.

Theory and Simulations

Simulation conditions have been taken into account to make the results valid for different illumination settings and NAs of projection lens. The simulation setting is an ArF lithographic tool, and the wavelength (λ) of the source is 193nm. The Alt-PSM measurement marks are imaged by a projection lens with variable NA. The values of 0.60 and 0.65 are chosen. It is convenient to normalize the spatial and frequency variables in units of λ/NA and NA/λ respectively. For example, the pitch value of an Alt-PSM is usually recorded as $p\lambda/NA$. The partial coherence (σ) variable range is set to be 0.3~ 0.8 typically, when the accuracy of the aberration measurement technique is evaluated, see (1-2). For this aberration measurement technique, the variation of sensitivity is larger and the accuracy of the coma measurement is higher. The coma sensitivity usually increases when the partial coherence decreases (7). In our simulations, the partial coherence is set to be the lowest to get the highest sensitivity of coma. The conventional illumination is used and σ is set to be 0.3.

Three groups of simulations were carried out. In the first group of simulations, only one nonzero Zernike aberration term $Z_7 = 0.05\lambda$ was introduced. Under this condition, a higher IPE was determined by a higher sensitivity to Z_7. In this group of simulations, *p* increased from 300nm to 700nm in 10nm steps. For each value of *p*, the IPE was simulated as a function of *pw*/*p*, and the maximum value of IPE was found to correspond to a specific *pw*/*p*. The values of the maximum IPE and the specific *pw*/*p* were recorded simultaneously. Then *p* increased into another value, while the IPE and *pw*/*p were* calculated repeatedly. The results of this group of simulation are shown in Fig. 1. The α-curve shows the value of *pw*/*p* as a function of *p*, when the *pw*/*p* causes the maximum IPE. The β-curve shows the maximum IPE as a function of *p*. From the β-curve, the maximum IPE increases with *p* when $p \le 1.92$. When $p = 1.92$, the maximum IPE reaches its extremum. When *p* is larger, it decreases. From the α-curve, the Alt-PSMs with *pw*=2*p*/3 introduce the maximum IPEs when$1.41 \le p \le 1.92$. An Alt-PSM with $p = 1.92$ and *pw*=2*p*/3 proves to be with the highest sensitivity to third-order x-coma (Z_7).

In the second group of simulations, only one nonzero Zernike aberration term $Z_{14} = 0.05\lambda$ was introduced. The IPEs caused by the fifth-order x-coma (Z_{14}) were negative, while the third-order x-coma (Z_7) caused positive ones, which were in an opposite direction. The same procedures of simulations were carried out as the first group. The results are shown in Fig. 2. The γ-curve shows the value of *pw*/*p* as a function of *p*,

when the *pw*/*p* causes the minimum IPE. The η-curve shows the minimum IPE as a function of *p*. An Alt-PSM with *p*=2.02 and *pw*=2*p*/3 proved to be with the highest sensitivity to fifth-order x-coma (Z_{14}). When *p* reduced from 2.02 to 1.92, the minimum IPE caused by Z_{14} increases no more than 2%. It can be accepted that an Alt-PSM with *p*=1.92 and *pw*=2*p*/3 proves to be with the highest sensitivity to coma aberration (Z_7 and Z_{14}).

In the third group of simulation, the correlation between IPE and *p* was studied. Two kinds of Alt-PSMs were imaged by the projection lens at different NAs, respectively. The ordinary Alt-PSMs were characterized with *pw*=1/2**p*, and the optimized ones were characterized with *pw*=2/3**p*. Each Alt-PSM was imaged at NA=0.60 and NA=0.65, respectively. IPE was simulated as a function of *p*, when different kinds of coma were introduced into the projection lens. They are Z_7 and Z_{14}. Both values of them were 0.05λ. Fig.3 and Fig.4 show IPE as a function of p when Z_7 was introduced. Fig.5 and Fig.6 show IPE as a function of *p* when Z_{14} was introduced. From Fig.3, IPE increases with *p* for the ordinary Alt-PSMs, when $p \le 1.18$. When $p = 1.18$, it reaches its extreme. Then it decreases when *p* increases. IPE increases with p for the optimized Alt-PSMs, when $p \le 1.92$. When $p = 1.92$, it reaches its extreme. Then it decreases when *p* increases. The similar results can be learned from Fig.4. Comparing Fig.3 and Fig.4, the simulation results were insensitive to the value of NA but to λ / NA. It is reasonable to use the normalized units. The similar analysis can be carried out with Fig.5 and Fig.6. It can be induced that the absolute value of IPE increases with p and reaches its maximum as *p* arrives at a certain value, then IPE decreases as p increases. The maximum IPE of the optimized Alt-PSM is larger than that of the ordinary one. When IPE reaches its extreme, the optimized Alt-PSM also appears with the larger pitch than the ordinary one.

In order to analyze the simulation results, we consider the transmission function t(x) of an Alt-PSM, which can be expressed as,

$$t(x) = \sum_{n=-\infty}^{+\infty} \delta(x - 2np) * \left[rect\left(\frac{x + p/2}{pw} \right) - rect\left(\frac{x - p/2}{pw} \right) \right] \quad n \in Z . \tag{1}$$

The spectrum of the Alt-PSM is the Fourier transformation of $t(x)$,

$$U(f_x) = \frac{j \cdot pw}{p} \sum_{-N}^{+N} \delta\left(f_x - \frac{n}{2p} \right) \sin c(pw \cdot f_x) \sin(\pi p f_x), \tag{2}$$

where f_x is the spatial frequency of the object spectrum in the x direction. $\pm N$ are the highest orders of diffraction light that can enter the lens pupil, and $\pm N/(2p)$ are their spatial frequencies. From Eq. (2), the diffraction light of even orders is missing. The ± 1-order diffraction light has the much larger amplitudes than that of other odd-order diffraction light. When $\lambda/[2(1-\sigma)NA] \le p \le 3\lambda/[2(\sigma+1)NA]$, only ± 1 orders diffraction light can enter the lens pupil (8-10), the image field is the sum of these two waves. The IPE introduced by ± 1 orders diffraction light can be estimated by the discrete gradient of the wavefront aberration when the coherent illumination is used (11),

Ex. 1016, Page 411

$$\Delta X_{1,-1} \propto -\frac{\phi_1 - \phi_{-1}}{f_1 - f_{-1}}, \tag{3}$$

where $f_1 = 1/(2p)$, $f_{-1} = -1/(2p)$ are spatial frequencies of +1 and -1 orders of the diffracted light, respectively. ϕ_1 and ϕ_{-1} are values of aberration function at f_1 and f_{-1}, respectively. Equipped with Eq.(3), $|\Delta X_{1,-1}|$ is found to be larger when f_1 and f_{-1} come nearer to the center of the pupil both for third-order x-coma $Z_7(3\rho^2 - 2)\rho\cos\theta$ and fifth-order x-coma $Z_{14}(10\rho^4 - 12\rho^2 + 3)\rho\cos\theta$. According to the analysis, the absolute value of IPE introduced by ±1 orders diffraction light is to increase with *p*. When $3\lambda/[2(\sigma+1)NA] \le p \le 5\lambda/[2(\sigma+1)NA]$, ±1 and ±3 orders diffraction light enters the pupil. The IPE comes to be the average of $\Delta X_{1,-1}$, $\Delta X_{3,-3}$, $\Delta X_{1,3}$, $\Delta X_{1,-3}$, $\Delta X_{-1,3}$ and $\Delta X_{-1,-3}$, when an ordinary Alt-PSM is imaged by the projection lens. The averaged IPE is lower than $\Delta X_{1,-1}$ and decreases when $p \ge 3\lambda/[2(\sigma+1)NA]$. But ±3 orders diffraction light proves to be modulated to zero for an Alt-PSM with *pw*=2*p*/3, which can be learned from Eq.(2). So, the IPE is only introduced by ±1 orders diffraction light for this kind of optimized Alt-PSM, when $3\lambda/[2(\sigma+1)NA] \le p \le 5\lambda/[2(\sigma+1)NA]$. The absolute value of the IPE increases with *p* continually, when $p \le 5\lambda/[2(\sigma+1)NA]$. Then it decreases when $p > 5\lambda/[2(\sigma+1)NA]$.

Considering with the third group of simulations, as shown in Fig.3, there is only ±1 orders diffraction light entering the pupil, when $p \le 1.18$. IPE increases with p both for the optimized and ordinary Alt-PSMs. When $1.18 \le p \le 1.92$, ±3 orders diffraction light enters the pupil. IPE begins to decreases for the ordinary Alt-PSM. But it still increases with *p* for the optimized one, because the ±3 orders diffraction light was modulated to zero. When $p > 1.92$, ±5 orders diffraction light enters the pupil and IPE of the optimized Alt-PSM also begins to decreases.

Considering with the first group of simulation, when $1.41 \le p \le 1.92$, ±1 and ±3 orders diffraction light enters the pupil. The Maximum IPE increases with *p* when ±3 orders diffraction light was modulated to zero by the optimized structure of *pw*=2*p*/3. When p>1.92, ±5 orders diffraction light enters the pupil, the Maximum IPE comes to be the average of $\Delta X_{1,-1}$, $\Delta X_{5,-5}$, $\Delta X_{1,5}$, $\Delta X_{1,-5}$, $\Delta X_{-1,5}$ and $\Delta X_{-1,-5}$. It is smaller than $\Delta X_{1,-1}$ and decreases as p increases. A similar analysis can be carried out for the second group of simulation. It is impossible to find a value of *pw*/*p*, which can modulate ±3 and ±5 orders diffraction light to zero simultaneously. A special value of *pw*/*p* can be found to make ±5 orders to be zero only, but ±3 orders may bring more reduction in IPE than ±5 orders, because their larger amplitude. Compared to other Alt-PSMs with different pitches and phase widths, an Alt-PSM with *p*=1.92, *pw*=2*p*/3 has the highest sensitivity to coma aberration (Z_7 and Z_{14}).

Accuracy Evaluation

In order to enhance the coma measurement accuracy, an optimized Alt-PSM with p=1.92 and pw=2p/3 was used as a measurement mark instead of an ordinary one with p=1.92 and pw=p/2. These two kinds of Alt-PSM marks with different structures are shown in Fig.7. To evaluate the accuracies of coma measurement for these two marks, the sensitivity variations of coma were calculated for them at multiple NA and σ settings by lithography simulator PROLITH. The accuracy of the coma measurement was estimated by (11),

$$MA \propto \frac{OA}{|S_{\max} - S_{\min}|}, \qquad (4)$$

where MA is the measurement accuracy of coma, and OA is the overlay accuracy of the lithography tools. $S_{\max}$ and $S_{\min}$ are the maximum and minimum of sensitivities to coma, respectively. In the simulation, the wavelength was 193nm. NA varies from 0.5 to 0.8. σ varies from 0.3 to 0.8. The overlay accuracy of the lithographic tool is assumed to be 1nm.The variation of sensitivities and the estimated accuracy are shown in table 1. From table 1, the measurement accuracies of Z_7 and Z_{14} have been enhanced by 15% and 31% respectively, by using the optimized Alt-PSM with pw=2p/3 as the measurement mark instead of the ordinary one with pw=p/2.

Conclusion

In conclusion, we have studied the correlation between the coma sensitivity of the Alt-PSM and the Alt-PSM's structure. For the ordinary Alt-PSMs, IPE increases with p before ± 3 orders diffraction light enters the pupil, then it reaches its extreme and begin to decrease. For the optimized ones, IPE increases with p before ± 5 orders diffraction light enters the pupil, then it reaches its extreme and begin to decrease. The structure of p=1.92 λ / NA, pw=2p/3 proves to introduce the highest sensitivity to coma. The measurement accuracies of Z_7 and Z_{14} have been enhanced by 15% and 31% respectively, by using the optimized Alt-PSM with pw=2p/3 as the measurement mark instead of the ordinary one with pw=p/2.

Acknowledgments

This work was supported by the Key Basic Research Program of Science and Technology Commission of Shanghai Municipality under Grant NO. 07JC14056. The authors would like to thank NERCLE (National Engineering Research Center for Lithographic Equipment, China) for the support of this work.

Ex. 1016, Page 413

References

1. F. Wang, X. Wang, M. Ma, D. Zhang, W. Shi and J. Hu, Appl. Opt. **45**, 281-287 (2006).
2. M. Ma, X. Wang, F. Wang, W. Shi and D. Zhang, Proc. SPIE **6150**, 615003 (2006).
3. E. Hendrickx, A.Colina, A. V. D. Hoff, J. Finders, G. Vandenberghe, J. Microlith.,Microfab.,Microsyst. **4**(3), 033006 (2005).
4. T.A. Brunner, IBM J. Res. Develop. **41**, 57-67 (1997).
5. Y. Ohsaki, T. Mori, S. Koga, M. Ando, K. Yamamoto, T. Tezuka, Y. Shiode, Proc. SPIE **6154**, 615424 (2006).
6. H. van der Laan, M. Dierichs, H. van Greevenbroek, E. McCoo, F. Stoffels, R. Pongers and R. Willekers, Proc. SPIE **4346**, 394-407 (2001).
7. Alfred K. Wong, Lars W. Liebmann, and Antoinette F. Molless, Proc. SPIE **4346**, 420-428 (2001)
8. H. Nomura and T. Sato, Appl. Opt. **38**, 2800-2807 (1999).
9. F. Wang, X. Wang and M. Ma, Appl. Opt. **45**, 6086-6093 (2006).
10. M. Ma, X. Wang and F. Wang, Appl. Opt. **45**, 8200-8208 (2006).
11. A. K. Wong, Proc. SPIE **4691**, 359-368 (2002).

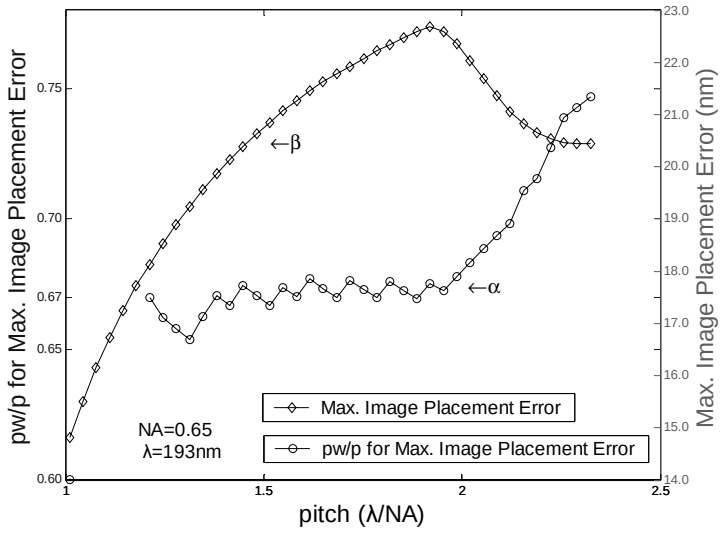


Fig. 1. The correlation between the Maximum IPEs and the structures of the Alt-PSMs. When only $Z_7 = 0.05\lambda$ is introduced into the projection lens, a Maximum IPE is found to correspond to a specific *pw*/*p for* each value of pitch.

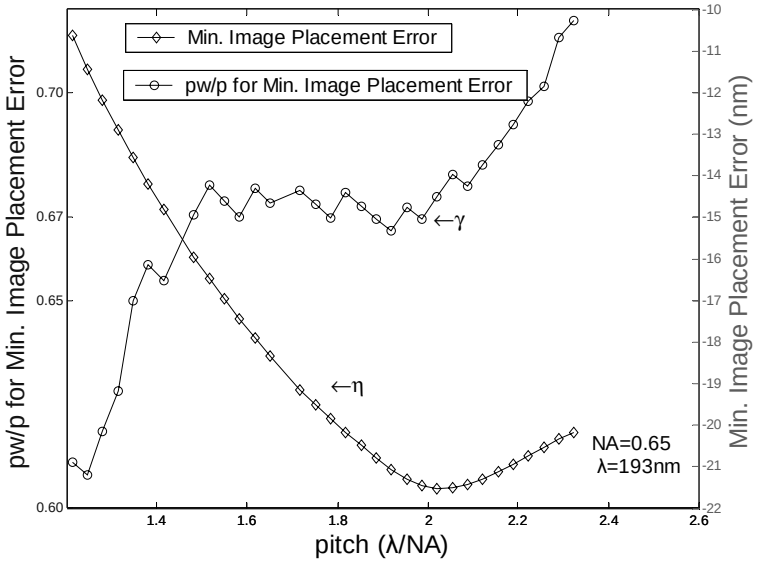


Fig. 2. The correlation between the Minimum IPEs and the structures of the Alt-PSMs. When only $Z_{14} = 0.05\lambda$ is introduced into the projection lens, a Minimum IPE is found to correspond to a specific *pw*/*p* for each value of pitch.

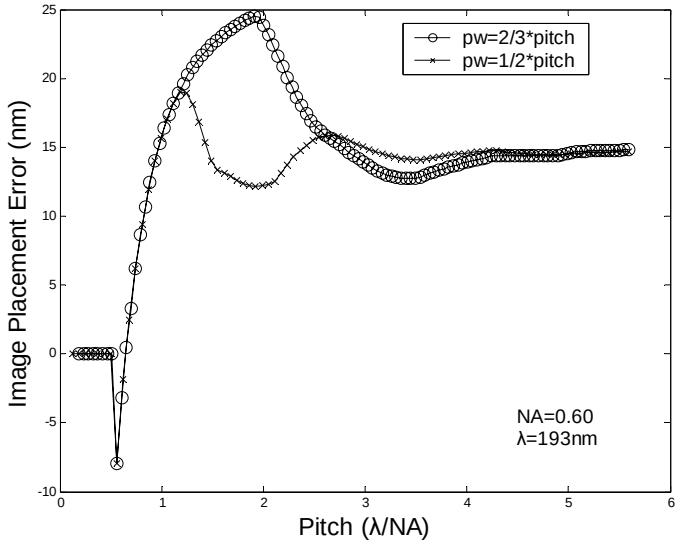


Fig. 3. The correlations between IPEs and values of pitches are shown for the ordinary Alt-PSMs and the optimized ones, when $Z_7 = 0.05\lambda$ is introduced into the projection lens and NA of it is 0.60.

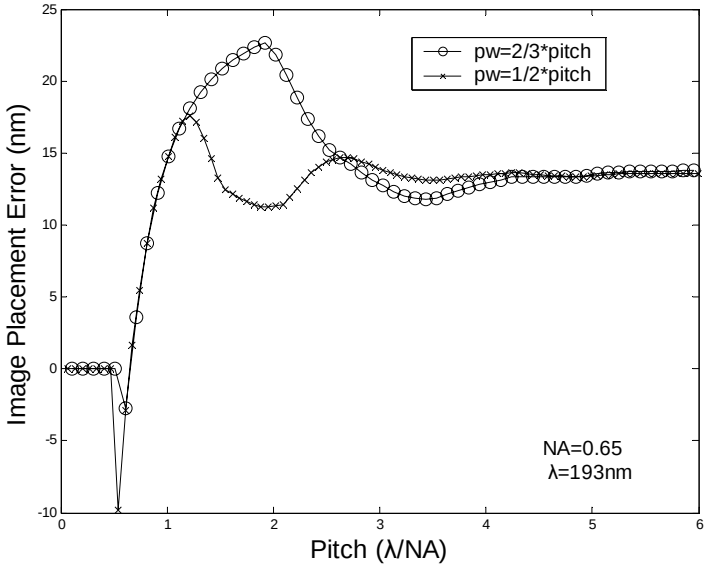


Fig. 4. The correlations between IPEs and values of pitches are shown for the ordinary Alt-PSMs and the optimized ones, when $Z_7 = 0.05\lambda$ is introduced into the projection lens and the NA of it is 0.65.

Ex. 1016, Page 416

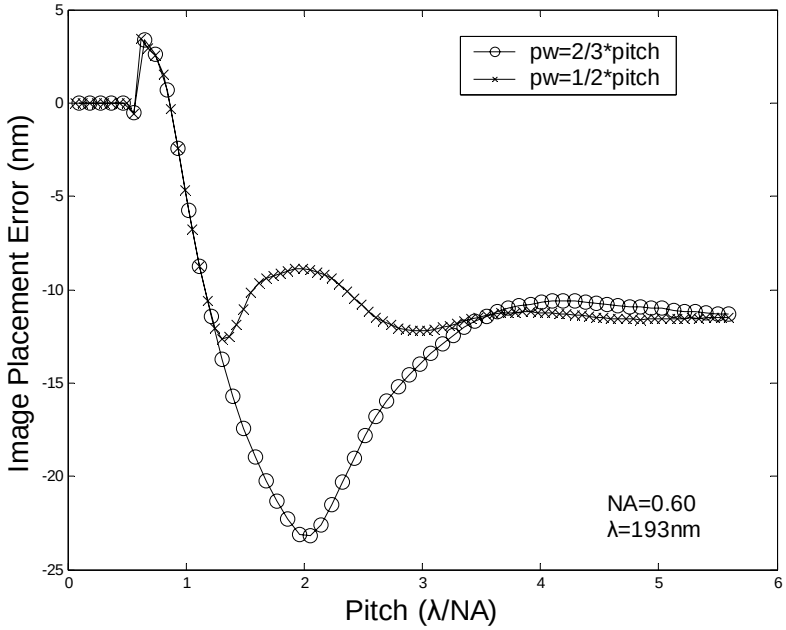


Fig. 5. The correlations between IPEs and values of pitches are shown for the ordinary Alt-PSMs and the optimized ones, when $Z_{14} = 0.05\lambda$ is introduced into the projection lens and NA of it is 0.60.

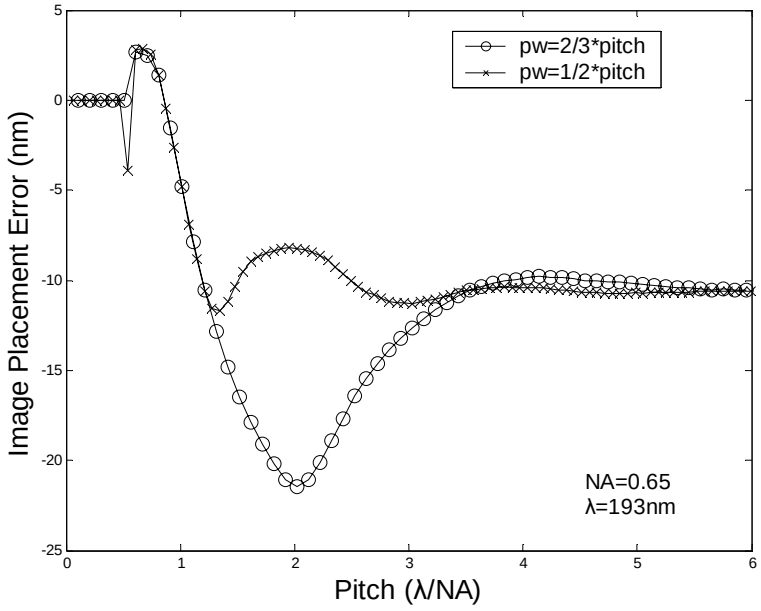


Fig. 6. The correlations between IPEs and values of pitches are shown for the ordinary Alt-PSMs and the optimized ones, when $Z_{14} = 0.05\lambda$ is introduced into the projection lens and NA of it is 0.65.

Ex. 1016, Page 417

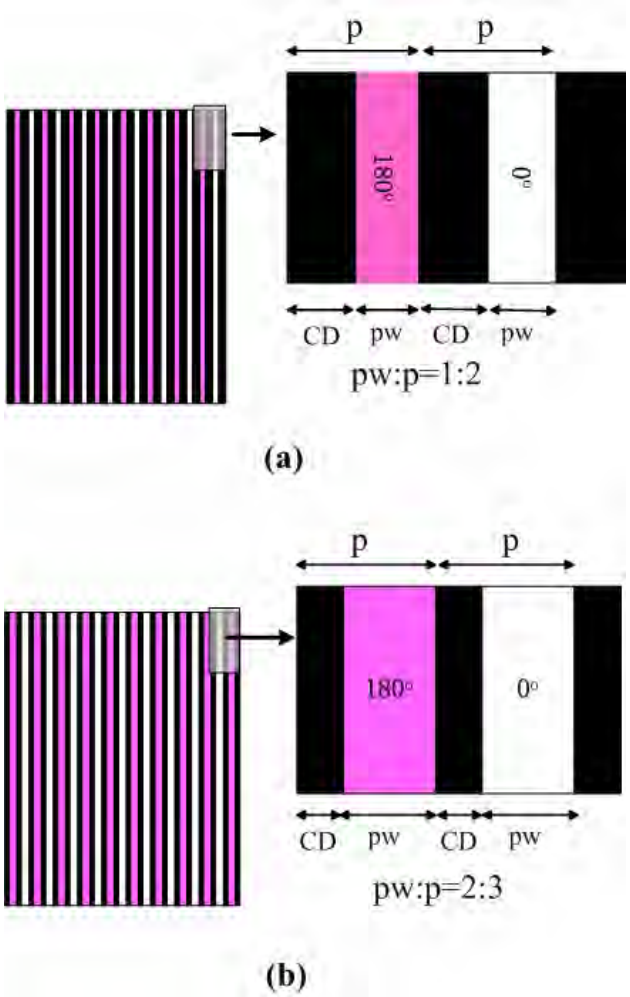


Fig. 7. (a) The ordinary Alt-PSM with *pw*:*p*=1:2; (b) The optimized Alt-PSM with *pw*:*p*=2:3.

Table I. Variation Ranges of Sensitivities and Accuracy of Coma Measurement

Zernike Coefficient	Mark Structure	Maximum Sensitivity	Minimum Sensitivity	Variation Range	Measurement Accuracy(nm)
Z_7	*pw*=2*p*/3	2.21	0.01	2.20	0.45
Z_7	*pw*=*p*/2	1.96	0.04	1.92	0.52
Z_{14}	*pw*=2*p*/3	0.35	-2.67	3.02	0.33
Z_{14}	*pw*=*p*/2	0.32	-1.97	2.29	0.44

ECS Transactions, 18 (1) 391-396 (2009)
10.1149/1.3096476 © The Electrochemical Society

EUV SOURCE SYSTEM DEVELOPMENT FOR 22NM GENERATION AND BEYOND

Benjamin Szu-Min Lin[a], David Brandt[b], and Nigel Farrar[b]

[a]Cymer Southeast Asia Ltd., 3F, No. 49, Lane 2, Kuang Fu-Rd., Sec. 2, Hsin-Chu, Taiwan, R.O.C.,
[b]Cymer Inc., 17075 Thornmint Ct., San Diego, CA 92127

Extreme ultra-violet (EUV) lithography is considered to be the most viable candidate to replace DUV to fulfill Moore's Law for 22nm generation and beyond. In this manuscript, we will show that LPP is the preferred choice for high volume manufacturing due to its scaling capability, enabled by high efficiency in both EUV light production and collection. The LPP EUV source development history and current status are described in this work, including EUV power progress, highly reliable Tin droplet generator, advanced debris mitigation technology and other technology advantages. Future EUV source roadmap provides a detailed schedule and technology requirements for high volume manufacturing EUV sources for the 22nm generation and beyond.

Introduction

EUV Lithography (EUVL) is the most promising technology as the next critical dimension imaging solution after 193 nm immersion lithography below the 32 nm node on the ITRS beginning in 2013. However, some memory devices, such as NAND Flash, may need EUVL technology as soon as 2011, and for development as early as 2009. High power and reliable EUV source has been categorized as high risk and ranked with other technologies requiring significant developments to enable the realization of EUVL. Photoresist performance parameters needing the greatest amount of development include sensitivity or speed, line edge roughness (LER), and line width roughness (LWR). Photoresist sensitivity and optical transmission drive EUV source power requirements within the usable bandwidth (BW) of 2 %. According to the joint requirements from scanner manufacturers an EUV power of >115W in 2 % BW at the intermediate focus (IF) is required for 5 mJ/cm^2 photoresist speed to enable >100 wph scanner throughput, and 180W in 2 % BW at IF is needed for 10 mJ/cm^2 resist. Photoresist sensitivities above 20 mJ/cm^2 could drive power requirements well above the 200W level, and therefore, a scalable EUV source architecture is needed to enable the evolution of EUV lithography during the life cycle of the technology. Low power discharge-produced-plasma (DPP) sources have been credited with the capability of producing 10W at IF and are expected to be scalable to the 30W - 50W range. Laser-produced-plasma (LPP) sources, on the other hand, are expected to deliver the necessary high power for high-volume manufacturing (HVM) scanners for the production of integrated circuits in the post-193 nm immersion era.

LPP EUV lithography light sources generate the required 13.5 nm radiation by depositing laser energy into a source element, such as xenon (Xe), tin (Sn) or lithium (Li), creating a highly ionized plasma with electron temperatures of several 10's of eV. The energetic radiation generated during de-excitation and recombination of these ions is emitted into all directions. It is collected with a normal-incidence mirror (collector), and focused to an intermediate point from where it is relayed to the scanner optics and ultimately to the

Ex. 1016, Page 419

wafer. The conversion efficiency (CE) of the laser energy into EUV energy is critical to meeting the required power levels. Previously, we have investigated and discussed several combinations of laser wavelength and source element with respect to CE and other aspects of light source integration[1]. In this report, LPP EUV power development with Sn droplet targets and CO2 laser radiation is described. Several key sub-system technology developments, Sn droplet generator, MLM collector, and debris mitigation, are also discussed. Finally, the future development roadmap illustrates the targets and requirements of HVM LPP EUV sources.

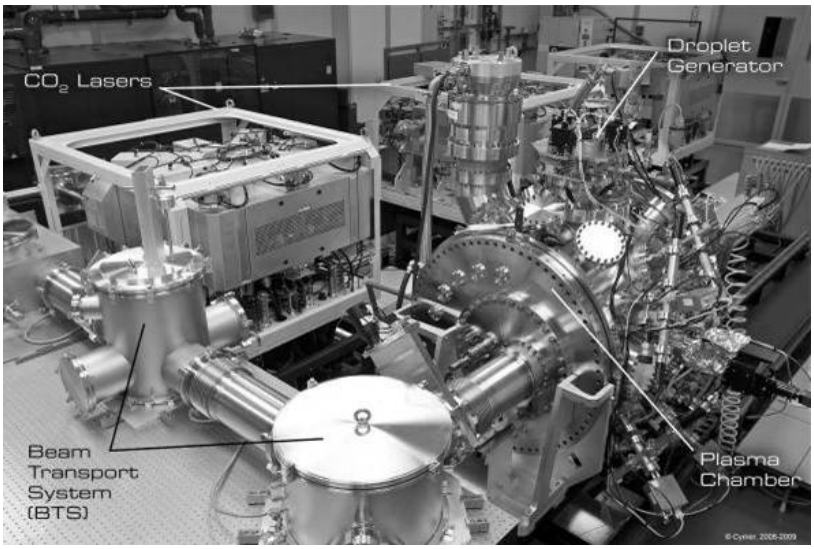


Figure 1. Integrated LPP EUV source system

LPP EUV Power Development Progress

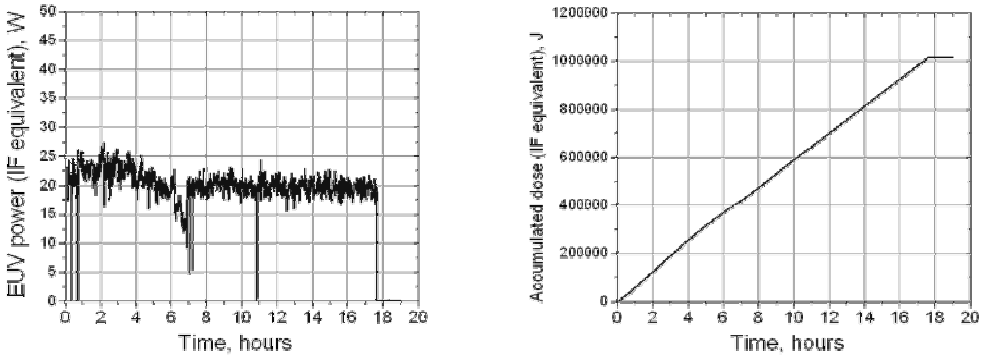
The laser produced plasma development system at Cymer was reconfigured to include a high-power high repetition rate CO2 laser, an axial-flow RF-pumped MOPA configuration with multiple stages of amplification[2], in the second quarter of 2006. The output power was measured to be on the order of 3 kW at 30 kHz during the installation phase. In the late fourth quarter of 2006 the CO2 laser system was upgraded to increase the output power to approximately 7 kW at 50 kHz. The CO2 laser pulse is amplified, shaped, and focused before entering the LPP vacuum chamber. The laser pulses are timed to strike an incoming liquid Sn droplet stream flowing from a target delivery system at a repetition rate of 50kHz. At the interaction region a plasma front with highly excited species is formed which radiates the desired 13.5 nm photons. A multilayer coated mirror tuned to reflect 13.5nm radiation collects 5sr of the back hemisphere and directs the power to IF at the output of the source chamber.

Starting from a prototype system level described two years ago[3], the main emphasis during the recent LPP development activities at Cymer has been on achieving substantial gains of the source output power by continuous improvement of laser intensity generated on the target droplet. The 100W burst power as well as 5W average power was achieved and confirmed the capability to operate at 3% conversion efficiency using liquid Sn droplet targets in Q4, 2007. To increase average power, solutions were required for thermal limitations in the laser, beam delivery and chamber components. In Q1, 2008, the

Ex. 1016, Page 420

average power was successfully increased to 35W, however, the operation duration was limited to only 2.5 seconds. In Q2, 2008, the operating time of the system was increased to hours (1.5 hours with 25W average power) due to improved thermal management capability. The above average power measurements were done without collector, and power was measured at the plasma and calculated at IF using standard assumptions of 5sr collection, 50% average reflectivity, and 90% transmission. Multiple 8 hour runs with a 320mm (1.6sr) MLM collector were then achieved in Q3, 2008. During these tests the burst power approached 50W, with 1ms burst duration at 8% duty cycle, as measured in the far field with a fluorescence converter and CCD camera. A photo diode with 1mm thick Zr foil was used to calibrate far field intensity to EUV photo diode looking at the plasma.

Recently the production systems capability has been increased to 400ms burst duration and 80% duty cycle, which is typical for scanner operation. Figures 2(a) and 2(b) show performance of the system when run under these conditions. Average power of approximately 20W 2% in-band at IF over an 18 hour continuous run is shown. The dose accumulation during this run was more then 1MJ, which is enough to process ~250 wafers assuming a dose of 10mJ/cm^2.



(a) (b)

Figure 2. (a) Average power ~ 20W at 80% duty cycle over 18 hours, and (b) accumulated dose over 18 hours.

Key LPP Sub-system Technology Developments

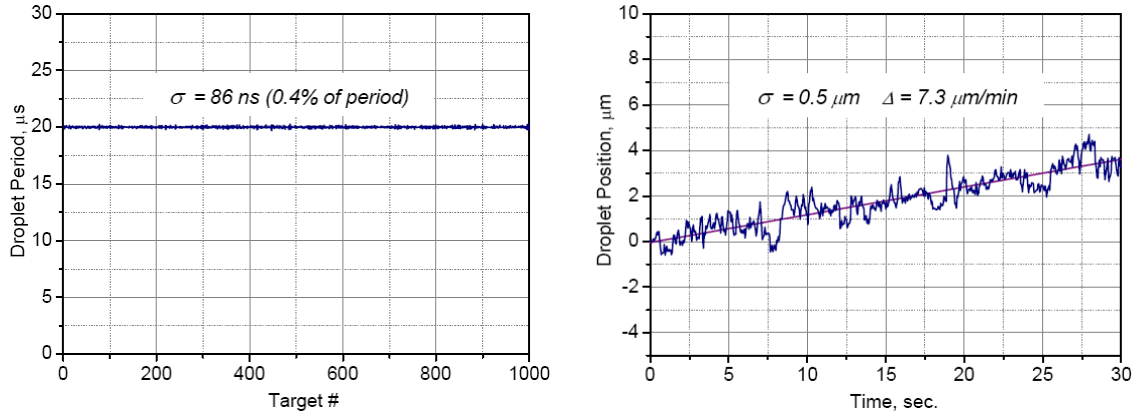
Besides obtaining higher average EUV power and longer operation duration, precise control of tin droplet generation, highly reflective MLM coating collector, and damage-free debris mitigation control should be developed for a robust integrated LPP EUV system. The development status of these three sub-systems is described as below.

Droplet generator

The main requirements for the droplet generator are the capability to produce tin droplets at a controllable frequency in the 20 – 200 kHz range with high droplet uniformity and stability and to provide reliable operation over long periods of time. The most recent droplet generator was designed to fulfill these criteria and represents a substantial performance improvement compared to earlier generators[4]. The stability of the droplets produced by the droplet generator is determined by adjusting the nozzle

parameters of the droplet generator (e.g., applied pressure and actuator driving voltage) and maintaining the overall mechanical stability of the system.

Figures 3(a) and 3(b) show 50µm diameter droplet timing stability and short term positioning stability, respectively, at a 50 kHz repetition rate. The standard deviation of droplet timing is 86ns (0.4% of the droplet period of 20µs). The standard deviation of short term droplet position is 0.5µm with a constant drift of 7.3µm/min. Even after testing over 500 hours, no degradation of timing or short term position stability can be found. Droplets as small as 30µm diameter are also meeting timing and short term positioning requirement. The Sn droplet diameter will be reduced to 10 microns in time for second generation EUV sources. The total tin consumption at high-frequency operation is estimated to be less than 1 liter/year/system under the assumptions of 85% utilization rate and 90% tool availability time.

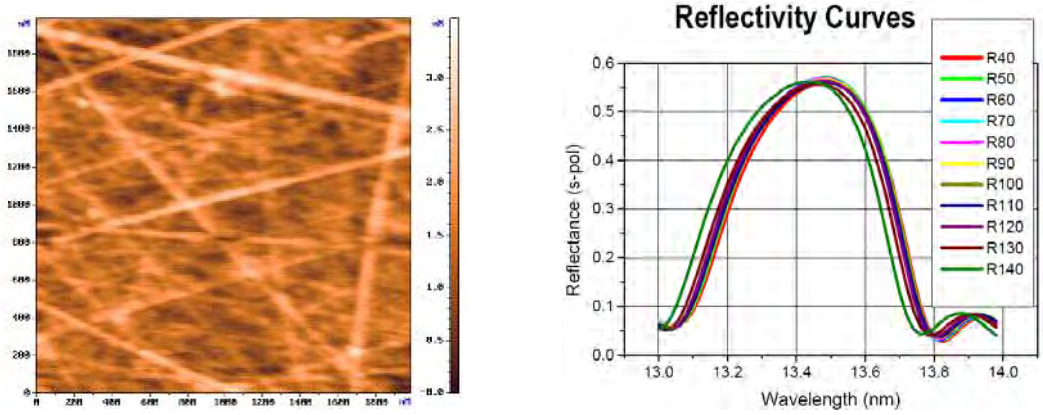


(a) (b)

Figure 3. (a) 50µm diameter droplet timing stability, and (b) short term positioning stability at a 50 kHz repetition rate.

MLM Collector

In our light-source configuration the laser beam is focused through a central opening of the collector mirror onto the droplet target for plasma generation. The EUV radiation emitted in the backwards direction is reflected at near-normal incidence by the ellipsoidal mirror and focused to the IF[5]. For demonstration of the light collection capabilities of our source, a 1.6sr sub-aperture version (300mm optical diameter) of the full HVM design with 5sr collection angle was implemented. After machining and super-polishing, the mirror surface was coated using a deposition tool and a technology based on DC magnetron sputtering compatible with large-area collector coating. A graded high-temperature-stable coating with interface-engineered multi-layers was applied to provide high EUV reflectivity at varying angle of incidence[6]. AFM measurement was done to make sure the surface was smooth enough to provide high reflectance. Figure 4(a) illustrates high spatial high frequency roughness of 0.452nm RMS within the area of 1.8µm x 1.9µm after MLM coating. The peak reflectance of the mirror was measured with synchrotron radiation (PTB, Berlin). Figure 4(b) displays the uniform reflectivity curves with peak reflectance of 57% for a series of measurement points at increasing distance from the mirror center.

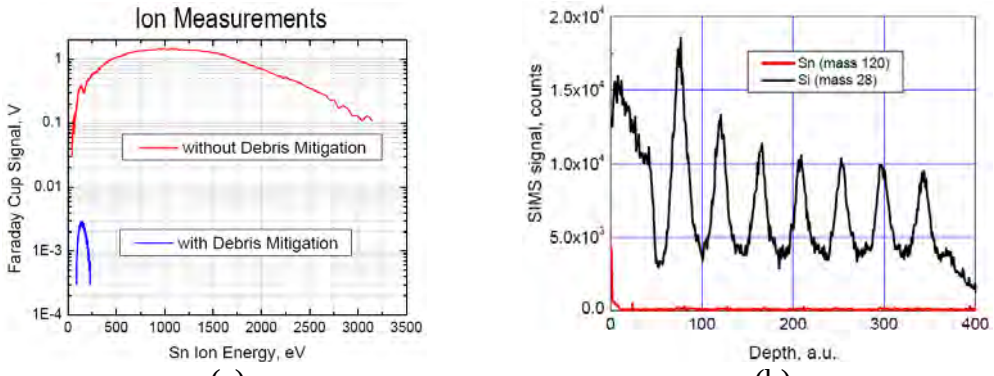


(a) (b)

Figure 4. (a) AFM measurement of MLM coating collector with 0.452 nm RMS within 1.8 x 1.9 um sampling area, and (b) uniform reflectivity curves for a series of measurement points at increasing distance from the mirror center.

Debris Mitigation

LPP EUV sources generate debris impacting on the surface of the collector in the form of high-energy ions, neutral atoms and clusters of target material. High-energy ions are the most hazardous among these three types of debris for the collector mirror coating. The impact of neutral atoms and clusters to damage the collector is rather smaller since most of the target material moves in a direction pointing away from the collector surface. The interaction of ions with energies of around a few kilo-electron volts with the surface results in erosion of the material of the MLM coating. An ion mitigation technique has been developed to provide a suppression of the ion flux by 4 orders of magnitude, shown in Figure 5(a). To determine the effect of the ion mitigation technique, a depth profiling analysis of multilayer samples, e.g. using secondary ion mass spectrometry (SIMS), is suitable if the layer number is not too high[3]. MLM samples with 8 bilayers and one capping layer were installed at the collector position and exposed to 2 hours of LPP source operation with IF equivalent in-band output power of 60 W at 10% duty cycle. After the exposure the samples were analyzed using depth profiling by SIMS (Millbrook Scientific Instruments PLC). The results are shown in Figure 5(b), and no ion erosion was found.



(a) (b)

Figure 5. (a) Ion measurement with and without debris mitigation, and (b) SIMS analysis of MLM samples after 2 hours of LPP exposure, and no ion erosion was found.

Future Development Roadmap

The LPP source roadmap is shown in Table 1. The pilot product is expected to meet requirements for preproduction or beta generation scanners in 2009. For this product the in-band EUV output power is targeting >100 W using a 11 kW CO2 laser system on Sn droplets with 3.0 % conversion efficiency. The normal incidence collector will use 5sr collection angle with a coating having an average EUV reflectivity approaching 60 %. Transmission losses due to obscurations, absorption, and debris mitigation techniques are projected to be less than 20 %. In line with the expected requirements for increased in-band EUV power at IF, HVM I and HVM II generation LPP EUV sources will be brought to market using 19 kW and >20 kW CO2 laser technology with moderate improvements in CE and collection efficiency. It is expected to deliver HVM I and HVM II generation LPP EUV sources in 2010/Q1 and 2011/Q3, respectively.

EUV Source Power Roadmap			
	Pilot	HVM I	HVM II
Drive laser power (kW)	**11**	**19**	**>20**
In-band CE (%)	3.0	3.5	4.0
Collection Efficiency (sr)	5	5.2	5.5
Collector Reflectivity (%)	>60	>60	>60
Optical Transmission (%)	80	85	90
Total EUV power at IF (W)	**>100**	**>200**	**>400**

Table 1. Projected LPP EUV Source Power Roadmap.

Conclusion

Laser-produced plasma has been shown to be a viable EUV source technology with scalability to meet requirements from leading scanner manufacturers and provide a path toward higher power as the lithography tools evolve over their life cycle. The pilot LPP source of >100W is expected to meet requirements for preproduction or beta generation scanners in 2009. HVM I and HVM II generation LPP EUV sources are expected to be delivered in 2010/Q1 and 2011/Q3, respectively.

References

1. B. A. M. Hansson, et., in: Proc. of SPIE Vol. 5751, Emerging Lithographic Technologies IX, R. S. Mackay, Ed., 248-259 (2005).
2. I. V. Fomenkov, et., in: Proc. of SPIE Vol. 6151, Emerging Lithographic Technologies X, M. J. Lercel, Ed., 61513X, (2006).
3. B. A. M. Hansson, et., in: Proc. of SPIE Vol. 6151, Emerging Lithographic Technologies X, M. J. Lercel, Ed., 61510R, (2006).
4. J. M. Algots, et., in: Proc. of SPIE Vol. 5751, Emerging Lithographic Technologies IX, R. S. Mackay, Ed., 248-259 (2005).
5. N. R. Böwering, et., in: Proc. of SPIE Vol. 6151, Emerging Lithographic Technologies X, M. J. Lercel, Ed., 61513R, (2006).
6. T. Feigl, et., in: Proc. of SPIE Vol. 6151, Emerging Lithographic Technologies X, M. J. Lercel, Ed., 61514A, (2006).

ECS Transactions, 18 (1) 397-404 (2009)
10.1149/1.3096477 © The Electrochemical Society

Novel Embedded Barrier Layer Materials for ArF Non-topcoat Immersion Applications

Deyan Wang, Chunyi Wu, Cheng Bai Xu, George Barclay, Peter Trefonas, and Min Xu

Rohm and Haas Electronic Materials
455 Forest Street, Marlborough, Massachusetts 01752, USA

With the decrease in pitch in the line/space patterning, micro-bridge defects have become the major defect in the immersion applications. As a result, reducing micro-bridge defect count is one of the key tasks for mass production of semiconductor devices using immersion lithography for both topcoat and non-topcoat processes. In this paper, we focus on the non-topcoat approach particularly the embedded barrier layer (EBL) technology (1-4). The advanced EBL materials discussed in this paper have demonstrated to be able to reduce total defect including micro-bridge defect count to the same level as that of a topcoat process. It was found that the developer solubility of the EBL materials in both bright and dark fields and the contrast of the EBL materials play important roles for reducing overall defectivity.

Introduction

The EBL technology has become very attractive in this industry since it not only simplifies the immersion lithography process with reduced material cost, but also serves as a means to control surface receding contact angle to allow fast scan speed for high productivity. This novel approach, however, encountered high defectivity in the earlier stage of the technology development. The high defectivity was mainly attributed to the hydrophobic nature of the EBL material which is enriched in the top portion of the photoresist film. The hydrophobic surface causes poor wetting of both developer and rinsing water. To address these issues, EBL materials were designed to be both photoswitchable and developer-switchable as discussed in our previous papers (3-4). Although the photoswitchable property of the EBL material makes the EBL material behave like a photoresist polymer, the EBL dissolution rate in the unexposed areas and the contrast of the EBL materials were found being among the key factors of affecting the overall defectivity to the photoresist that contains the EBL material.

To further reduce the impact of an EBL material to the lithographic performance of an existing photoresist formulation, our advanced EBL materials are designed to be more effective in migrating to the resist surface during the spin coating of the photoresist. Thus the effective amount of an EBL material in a photoresist formulation is within 1 to 5%, relative to the total solid content of the photoresist. More hydrophobic EBL materials in a photoresist formulation often cause a poorer lithographic performance especially higher defectivity.
In this paper, we focus on the defectivity of EBL containing photoresist to discuss the impact of EBL developer dissolution rate in the dark field, and EBL contrast to the overall lithographic defectivity and micro-bridge defect count.

Ex. 1016, Page 425

The EBL Dark Field Developer Solubility and Defectivity

Although the photoswitchable property of an EBL material ensures the developer solubility of the EBL material in the exposed areas like a photoresist polymer, the dark field developer solubility of an EBL material was found to be crucial in controlling the defectivity of a photoresist formulation. Table 1 shows the comparison of normalized total defect count and micro-bridge defect count for 48nm 1:1 L/S between the resist formulations containing EBL with and without developer dissolution rate in the dark field. The test was performed with ASML 1700i using 500mm/sec scan speed. As seen, the total defect count and micro-bridge defect count drop approximately 5 fold for an EBL material without developer dissolution to with developer dissolution rate.

Sample	Normalized Total Defect Count	Normalized Micro-bridge Defect Count
EBL without developer dissolution rate in dark field	15.3	5.33
EBL with developer dissolution rate in dark field	3.25	1

Table 1. Normalized defect count comparison between with and without developer dissolution rate of an EBL material. The micro-bridge defect count of the EBL having developer dissolution rate is normalized to one.

Further defect tests were conducted with photoresist formulations that contain EBL materials with different developer dissolution rates in dark field. The different developer dissolution rates of the two EBL materials are depicted in Figure 1.

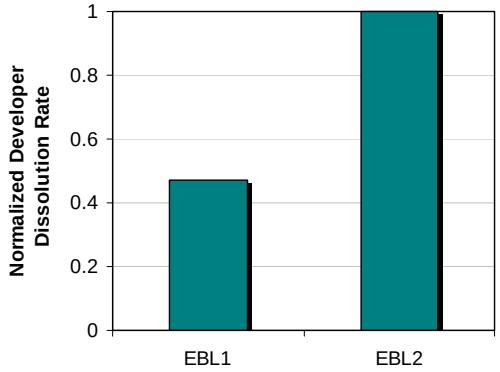


Fig. 1. Developer dissolution rates of EBL1 and EBL2 with that of EBL2 normalized to one.

The defect test was conducted on ASML 1100 for 75nm 1:1 L/S using dipole illumination. Shown in Figures 2 and 3 are the comparison of the normalized total defect count and micro-bridge defect count.

Ex. 1016, Page 426

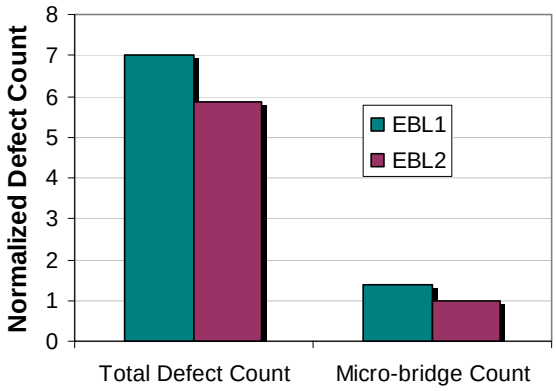


Fig. 2. Total defect count and micro-bridge defect count comparison between EBL1 and EBL2 with the micro-bridge defect count of EBL2 normalized to one.

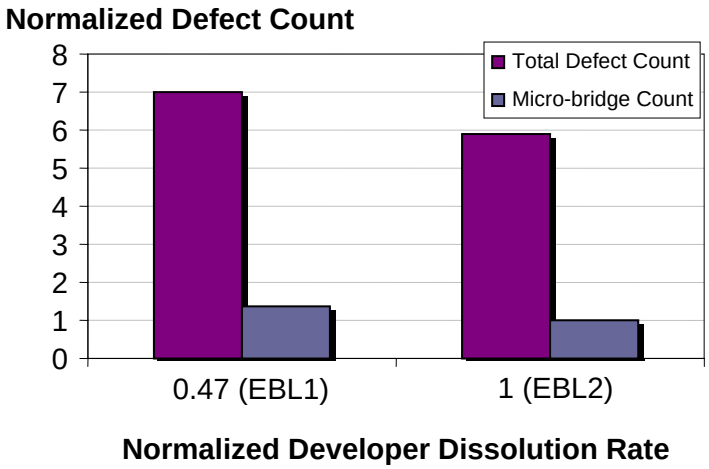


Fig. 3. Total defect count and micro-bridge defect count as a function of developer dissolution rate of the EBL materials with the micro-bridge defect count normalized to one for the higher developer dissolution rate.

The EBL with a higher developer dissolution rate has slightly lower defect count (both total defect count and micro-bridge defect count). However, a 50% decrease in developer dissolution rate of the EBL materials only causes a slight increase of the defect count. The defect count in dark field showed no significant difference between these two EBL materials.

To summarize, the EBL material with dark field developer solubility is able to reduce the defectivity dramatically. However, a further increase in developer dissolution rate of the EBL material only leads to a slight decrease of the defect count.

EBL Contrast and Performance

As discussed in our previous publications (1, 3-4) the novel EBL materials that we developed are photoswitchable by incorporating acid labile functional groups into the EBL polymers. The acid labile functional groups make the EBL materials become developer soluble or more soluble in the exposed areas upon post exposure bake (PEB).

Ex. 1016, Page 427

When a small amount of EBL was added in a photoresist, the EBL on the photoresist film surface and in the gradient region next to surface may impact the photoresist performance particularly the overall defectivity. The impact is partially due to the difference in dissolution contrast between the EBL materials and the photoresist polymers.

In this section of the experiments, two pairs of EBL polymers were studied. In each pair, the EBL polymers are the same in compositions but have a distinguishable difference in contrast. To illustrate the contrast difference between two EBL materials, the EBL materials were used as resist polymer and formulated with identical PAG, quencher and their concentrations. PGMEA was used as a solvent. Both formulations were filtered with 0.2µm TFPE filters prior to the contrast curve test. The formulation was coated with the same soft baking temperature for 1200A film thickness on the same antireflective coating, and then the same PEB temperature was applied after the 193nm wavelength exposure. The contrast curves of the two EBL materials A and B are shown in Figure 4. It is clearly seen that EBL B has lower contrast than EBL A.

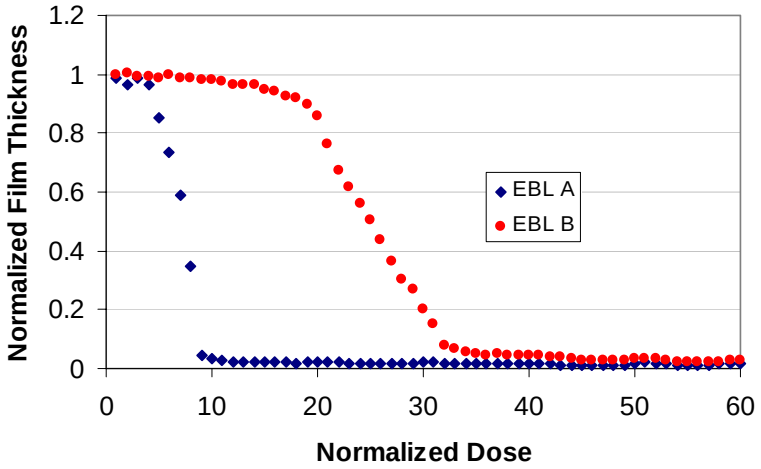


Fig. 4. Contrast curves of EBL materials A and B with B being lower in contrast. Both EBL materials were tested under the identical conditions.

To compare the effect of these two different EBL materials as barrier layer to the performance of a photoresist, an equal amount of EBL A and B was added in the same photoresist formulation to form two new formulations. The surface static and dynamic contact angles of the two resists were measured afterwards, and it was found that both samples have essentially the same static and receding angles (~73° for receding angles).

The lithographic tests were conducted with ASML 1100 scanner under the identical processing conditions for 75 nm 1:1 L/S patterns using dipole illumination. Shown in Figure 5 are the energy latitude (EL) images of the two resists containing EBL A and B. No significant difference between the two samples was observed.

Ex. 1016, Page 428

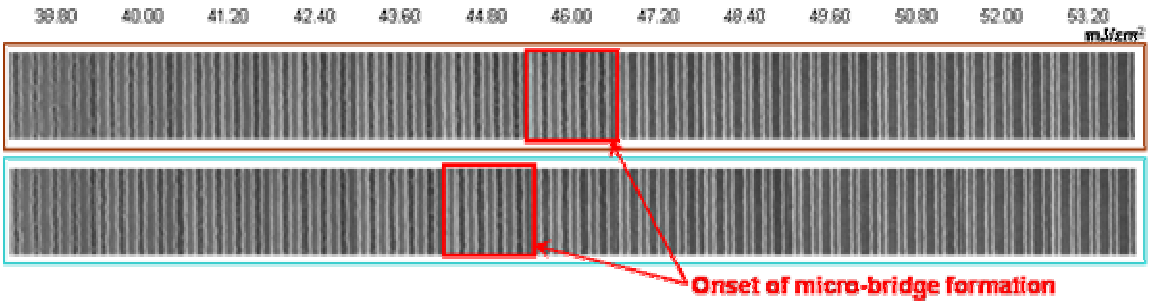


Fig. 5. EL images of the photoresist samples containing EBL A (higher contrast, the bottom row) and EBL B (lower contrast, the top row)

The defect test was carried out with an in-house small dispersing system. The normalized defect counts of the resists with these two EBL materials are given in Table 2. The photoresist containing EBL A (higher contrast) results in significant lower total defect count and micro-bridge defect count than that of the resist containing EBL B (low contrast). More than three folds of micro-bridge defect reduction were achieved by using the higher contrast EBL A.

Sample	Normalized Total Defect Count	Normalized Micro-bridge Defect Count
Photoresist Containing EBL A	2.25	1
Photoresist Containing EBL B	5.21	3.21

Table 2. Normalized defect count comparison between EBL A (higher contrast) and EBL B (lower contrast) for 75 nm 1:1 L/S using ASML 1100 scanner with dipole illumination

Similar study was conducted with another pair of EBL materials (EBL C and D) with the same composition. EBL C has higher contrast than EBL D in this pair. EBL C and EBL D differ from EBL A and EBL B in the polymer chain structures, and it was found that the contrast effect is more prominent in the pair containing EBL C and EBL D. Shown in Figure 6 are the exposure latitude images of the photoresist samples containing EBL C and EBL D, respectively. It is easily discerned that the resist with EBL D (lower contrast) suffers from heavy bridge defect, residual defect and higher line edge roughness (LER).

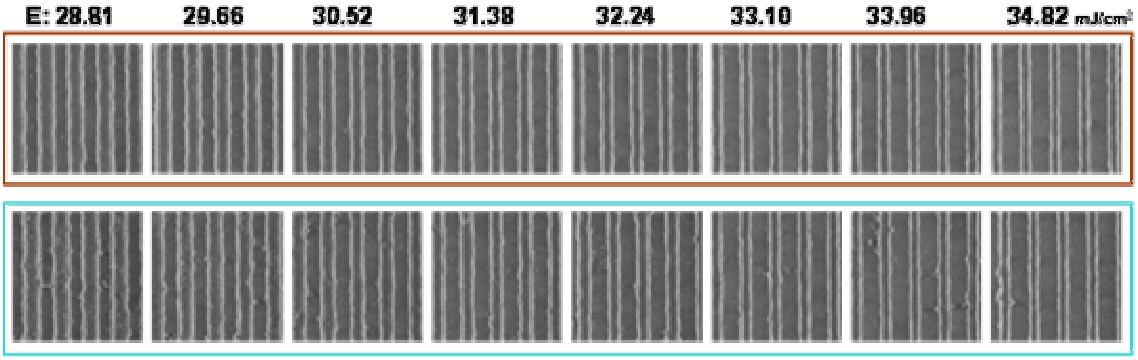


Fig. 6. EL image comparison between EBL C (higher contrast, the top row) and EBL D (lower contrast, the bottom row), for 75 nm, 1:1 L/S using ASML 1100 scanner with dipole illumination

The two pairs of EBL materials discussed in this section represent two different types of EBL polymer backbones. Regardless the type of the polymer chain structures, the EBL materials with higher contrast, in general, lead to fewer defects.

Discussions

Even with the photoswitchable functional groups built into the EBL polymer, the dark field developer dissolution rate still plays a very important role in reducing overall defectivity particularly the micro-bridge defect count. The hydrophobic EBL material enriched on the resist surface in the darker field is believed to be the root cause of the higher defectivity. EBL materials with a developer dissolution rate allow the developer to remove the thin layer of EBL (<10nm) and create a surface more like the original photoresist surface without the EBL. Therefore, EBL materials with developer solubility lead to lower defectivity.

It is worthwhile to note that the functional groups incorporated into an EBL material not only affect the EBL barrier property, namely the effectiveness for the EBL materials to migrate to the top surface to form a decent leaching barrier and to provide high receding angles with a lower EBL loading, but also change the optical behavior of the EBL materials. Therefore, proper functional groups need to be selected to balance between EBL physical property and lithographic performance especially the overall defectivity.

EBL 2 as discussed previously is among the successful EBL materials that we developed to balance between higher receding angle and lower defectivity while maintaining the good lithographic performance of the resists that contain the EBL materials. Shown in Table 3 are the lithographic test results for EBL 2 in a resist along with the resist using a top coat as a reference.

Sample	Esize @ Target	MEF	EL by +/- 10% of target CD	LER	LWR	PCM	Total defect count*	µ-bridge defect count *
Resist with Top Coat	29.7	1.43	27.7	3.9	5.0	39.3	3.7	2.6
Resist with EBL 2	30.1	1.40	34.8	3.8	4.6	39.4	3.2	1

Table 3. 45nm 1:1 L/S obtained with ASML 1900i using dipole illumination, NA 1.3, outer/inner sigma 0.97/0.82.
* Normalized defect count obtained with ASML 1700i. The micro-defect count for Resist with EBL 2 was normalized to one.

The top-down and the cross-section images are shown in Figure 7.

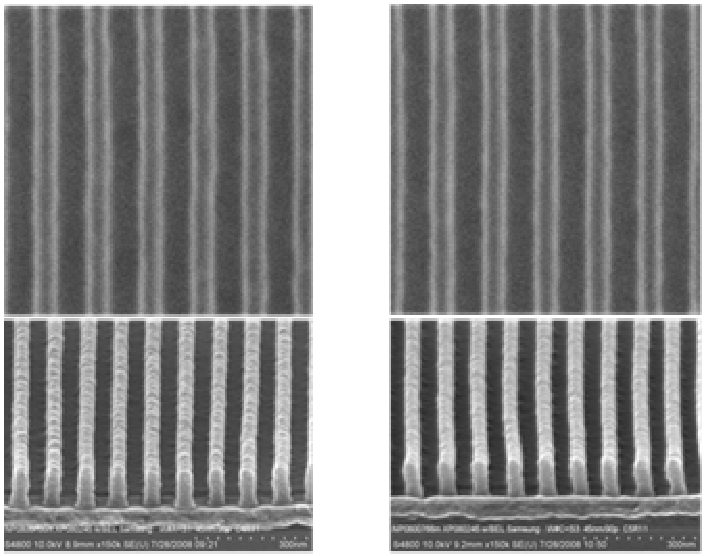


Fig. 7. 45nm 1:1 L/S obtained with ASML 1900i using dipole illumination, NA 1.3, outer/inner sigma 0.97/0.82. On the left are the resist with an immersion top coat, and on the right the resist with EBL 2.

Conclusions

By introducing dark field developer solubility and enhancing the contrast of the EBL materials, we have successfully demonstrated that our novel EBL materials are able to increase resist receding angle for higher scan speed, and to reduce defectivity as low as that of a top coat process while maintaining the good lithographic properties of the resist and keeping low PAG quencher base leaching as we previously reported. We believe that the EBL technology is going to drive the immersion applications from topcoat to a non-topcoat process.

References

1. Rohm and Haas Electronic Materials, US patent application US20060087286 A1, European patent application EP1720072 A1 and Japan patent application JP2006309245 A.

2. Deyan Wang, Stefan Caporale, Cecily Andes, Kap-Soo Cheon, Cheng bai Xu, Peter Trefonas and George Barclay, "Design Consideration for Immersion 193: Embedded Barrier Layer and Pattern Collapse Margin", Journal of Photopolymer Science and technology, 687-696, Volume 20, Number 5, 2007.

3. Deyan Wang, Cheng bai Xu, Stefan Caporale and Peter Trefonas, "Creating 193nm immersion resists with embedded top barriers", Solid State Technology, Sept. 2007.

4. Steve Wu, Aroma Tseng, Bill Lin, Chun Chi Yu, Bo-Jou Lu, Wen-Sgiang Liao, Deyan Wang, Vaishali Vohra, Cheng Bai Xu, Stefan Caporale and George Barclay, "Non-topcoat Resist Design for Immersion Process at 32-nm Node", Proc. Of SPIE Vol. 6923, 692307-1 to -14

ECS Transactions, 18 (1) 405-408 (2009)
10.1149/1.3096478 © The Electrochemical Society

Silicon-Containing Coupler Polymers Used For DUV Photoresists

Ruicheng Ran

SuZhou CitiTech Advanced Material Inc. China

A new series of photoresists formulated by silicon-containing coupling agents copolymer is introduced to use for DUV microlithography. Those photoresists improved adhesion to substrates, contrast of pattern, lateral resolution and etching-resistance in lithography processing.

It is well known that silicon is a ground base of IC industry for a long time. Silicon-containing materials have been widespread used in IC industry, such as crystalline, polycrystalline, amorphous, and oxidative silicon on wafer, and promoter, CMP, low-*k* material, BARC and photoresists in microlithography process. In the past years, we have developed a new series of silicon-containing photoresists for DUV microlithography. Generally, the silicon-containing DUV photoresists used for the microlithography processing contain some electronic copolymers with silicon-containing composition, witch is applied as functional groups. These silicon-containing compositions are some kinds of silicon coupling agents (couplers). General, couplers contain two functional parts in their molecular. One part of the molecular can be reacted with inorganic materials such as silicon on surface of wafer, etc. Another part of the molecular is able to be polymerized with vinyl-type monomers or to be reacted with organic materials such as polymers. For instance, the coupler is an acrylate monomer with siloxane group (silicon alkyloxy group), $CH_2=CR_a\text{-}COO(CH_2)_n\text{-}Si(OR_b)_3$, where R_a=-H, $-CH_3$; R_b=-$C_{1\text{-}4}$ alkyl, n=1-8. This coupler is very easy to carry out a free radical polymerization with many vinyl monomers to form a copolymer, witch can be used as film formation material in a DUV photoresist, as shown below. In this copolymer the silicon-containing coupling agent become side-chain on the copolymer main chain, e.g. the coupling agent is combined upon the polymer. After the copolymer is formulated with photoacid generator (PAG), amine, leveling agent and solvent to form a photoresist, the functional behavior of coupling agent will remain and help photoresist to improve its adhesion, resolution and etch-resistance properties.

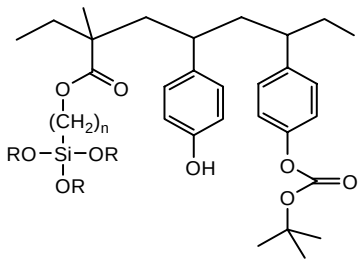


Fig. 1. Schematic for the silicon containing coupler polymer in a typical DUV photoresist

Ex. 1016, Page 433

1. Improvement of Photoresist Adhesion

In general, substrates (virgin wafer or wafer with fresh thermal oxide) need clearing or/and promotion before performing lithography process. After SiO_2-etching with HF, the photoresist adhesion strongly depends on the completeness of SiO_2 removal: with the SiO_2 completely removed, the now H-passivated Si-surface reveals a very good adhesion, while remaining SiO_2 causes a very bad and non-reproducible adhesion. Because on SiO_2 and many metals surfaces, the air humidity forms polaric OH-bonds, such substrates are hydrophilic (good adhesion of water), but reveal a bad wetting and adhesion of the unpolaric photoresist. For improving the adhesion, some adhesion promoters such as HMDS were used to treat the substrates making the hydrophilic surface hydrophobic caused by $-Si(CH_3)_3$ of HMDS chemically bonding to oxygen of SiO_2 as releasing ammonia (NH_3). However, our new photoresist formulated by the copolymers contained the silicon-couplers showed very good adhesion on substrates with either Si-surface or SiO_2-surface. Because the hydrophilic side of the Si-containing coupler (such as $-Si(OCH_3)_3$), this photoresist showed good adhesion capability with any untreated parts of SiO_2-surface on the substrates if the promotion went imperfectly. In many cases, during soft baking the hydrophilic side of the coupler may occurred hydrolysis with trace moisture in the resist to produce some partly hydrolysis products (such as $-Si(OH)(OCH_3)_2$, $-Si(OH)_2(OCH_3)$…) which will increase its hydrophilic and show better adhesion with SiO_2-surface.

In addition, in the photoresist Si-containing couplers always self-assemble and concentrate the Si-containing functional group side upon the surface of the resist due to its very low cohesive energy density (δ) and surface tension. This will help the coupler to improve resist adhesion a lot with even a very small amount in the copolymer.

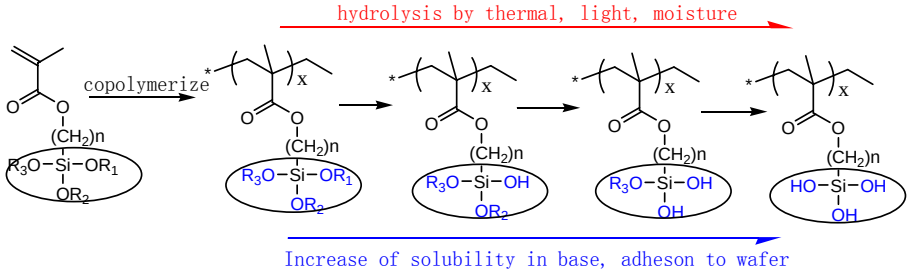


Fig. 2. Schematic of the hydrolysis process, which will improve solubility of the photoresist in base solutions.

2. Improvement of Photoresist Resolution

In lithography process, resolution depends upon many matters such as exposure machine, mask, technical processing and materials. Here we discuss only about some influences of photoresist on the resolution. For a chemically amplified resist (CAR), resolution depends mainly on improvement of its contrast to give a clear pattern.

The photoresist formulated with Si-containing couplers coplymers can get better contrast in the lithography process. After exposure with suitable dose and expose time, the resist film will be heated on a hotplate (PEB processing for CAR), and then developed with a basic developer to give a pattern. After exposure photoacid generator (PAG) produced strong acid (H^+) in exposed area of the resist. During the PEB process, the acid not only reacts with acid-labile protecting groups in the resist, but also reacts with the Si-containing couplers to make the resist more base soluble in developer such as TMHA ($(CH_3)_4N^+OH^-$) aqueous solution.

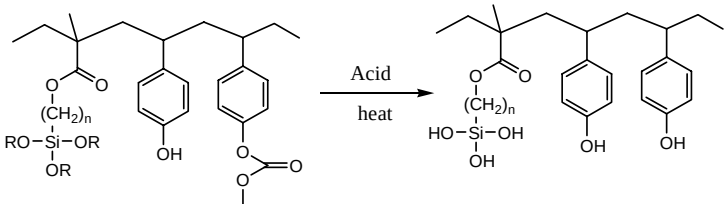


Fig. 3. Schematic of the reaction of the Si-containing coupler with acid and heat.

At the same time, the Si-containing couplers stay at unchanged and stable in unexposed area wing to without acid been produced by PAG.. So, the solubility different of the resist film in basic developer was enlarged between the exposed and the unexposed area. This will increase the contrast of pattern and help get a sharp structure profile. Therefore, this will improve the resolution, specially the lateral resolution of those photoresists.

3. Improvement of Etching Resistance

There are many deferent etchants for deferent etching processing and for deferent materials to be etched. In the etching processing, most importance properties for resists with good etching-resistance are good stability to the etching agents (etchants) especially to wet chemical etchants, and good adhesion to substrates. Generally, the unexposed photoresist film is soluble for organic solvents, but is insoluble for many etchants such as inorganic acid, HF, HCl, HNO_3, H_2SO_4, …The hard bake performed after development intends to increase the thermal, chemical and physical stability of developed resist structures for subsequent processes e.g. electro-plating, wet-chemical and dry etching.

The photoresist formulated based on the Si-containing couplers coplymers showed good etching resistance properties. Those resists in unexposed area are very stable to acidic etchants, and it will become more stable after hard baking because a light cross-linking reaction will occur between two Si-containing coupler sides at high temperature during hard bake processing even without acid but moisture. The cross-linking reaction is a multi-stage reaction. Some reaction stages are showed below, where R is an alkyl group with 1-4 carbon. These reactions are complicated but can be controlled by moisture, temperature and bake time.

Ex. 1016, Page 435

$$\text{--Si(OR)}_3 + \text{H}_2\text{O} \rightarrow \text{--Si(OH)(OR)}_2 + \text{ROH} \quad \text{(alcohol)}$$
$$2\ \text{--Si(OH)(OR)}_2 \rightarrow \text{--(RO)}_2\text{Si-O-Si(OH)(OR)--} + \text{ROH}$$
$$\text{--Si(OH)(OR)}_2 + \text{-Si(OR)}_3 \rightarrow \text{--(RO)}_2\text{Si-O-Si(OR)}_2\text{--} + \text{ROH}$$

On surface of the developed resist structure, the above cross-linking reaction occurs easily due to acid presence during etching. This will make surface more stable to etching agents.

In addition, the good adhesion to substrates will avoid many etching problems such as lifting of small structures, under etching or large-scale resist peeling as showing below. It is evident that all of those problems during etching process are related to the adhesion between resist and substrates.

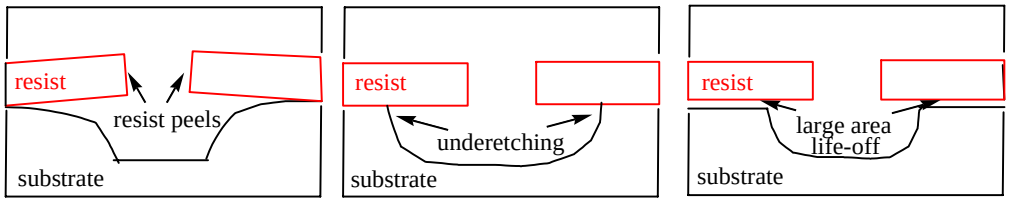


Fig. 4. Schematic of various problems during and after developing process originated from bad adhesion.

In conclusion, the DUV photoresists formulated with Si-containing coupler copolymers provide some very good properties in lithography processing, such as adhesion, resolution and etching-resistance. These resists would be commercialized soon.

References

1. Jean T-S, Song T-T., etc. J Photopolym Sci. Technol., 14(3), 503-506 (2001)
2. Kwark Young-J, Bravo Vasquez, etc. SPIE, Volume 5039, pp. 1204-1211 (2003).
3. Jin-Baek Kim, Ramakrishnan Ganesan, etc. *J. Mater. Chem.*, 2006, 16, 3448 – 3451
4. M. Toriumi, H. Shiraishi, etc. J. Electrochem. Soc., Vol. 134(4), pp. 936-939 (1987)
5. E.S. Moyer, J. Bremmer, etc., Semiconductor International , Sept. 2007

Ex. 1016, Page 436

ECS Transactions, 18 (1) 409-417 (2009)
10.1149/1.3096479 © The Electrochemical Society

A Cost-effective Spin on Sidewall Material Alternative to the CVD Sidewall Process

Bang-Ching HO*, Daisuke MARUYAMA, Rikimaru SAKAMOTO,
Yasushi SAKAIDA, Hiroaki YGUCHI,

Semiconductor Materials Research Department,
Electronic Materials Research laboratories,
Nissan Chemical Industries, Ltd.
635 Sasakura, Fuchu-machi, Toyama 939-2792 Japan

hobangching@nissanchem.co.jp

193nm immersion and Hyper NA lithography are used at 45nm and beyond. The next generation of lithography will use a new technology such as Double Pattering, EUV or EB. Double patterning is one of the currently acceptable technologies. Three common double pattern techniques are Litho-Etch-Litho-Etch (LELE), freezing, and sidewall (spacer) process. From a technical standpoint LELE is a very promising process, except for the second litho alignment. However, the cost of ownership will be very high because LELE will cost about twice as much as the current single litho patterning process. In order to build up a suitable double patterning technique, many device makers are developing unique processes. Two of these processes are freezing and sidewall. Flash memory makers are diligently investigating the sidewall process by CVD. This is because of the lack of a second litho alignment step, even with its high cost. The high cost of the CVD process can be reduced if a spin on material is used.
One of the goals of this paper is to reduce the cost of ownership by using spin on coatings for the sidewall process. Currently we are investigating this approach to control the sidewall width, profile and other properties.

Introduction

Double patterning is one of the currently acceptable technologies for next generation. Especially sidewall process was focused by flash memory company. However, it's still a higher cost process than conventional one because introduction of CVD process. In order to improve it, we would like to investigate spin-on type sidewall material. This target material was coated on resist pattern and was aimed for chemical reaction with resist acid for getting sidewall.

In this paper, we would like to discuss about controlling the spin on sidewall material.
1. Investigate of Spin on sidewall Process
    (1) Baking & Irradiation
    (2) Developer
2. Study of controlling sidewall CD
    Exposure, post bake condition
3. Research in to sidewall reaction

Ex. 1016, Page 437

Experimental

1. Measurement of Solvent Resistance

A stripping test was performed to determine the resistance of resists to solvents of spin on sidewall. In this procedure, the photo resist was spin-coated on silicon wafer and bake. The photo resist thickness was then measured. Spin on sidewall solvent, for example, top coat solvents was dispensed on silicon wafer on which the photo resist was coated, then kept for 60sec, followed by spin drying at 3500 rpm for 30sec to remove the solvent. The film was then baked on hot plate at 100 degree C for 30sec. The photo resist thickness was again measured. The amount of solvent resistance was determined to be the difference between the initial and final photo resist thickness.

2. Preparation of Resist Pattern

Patterned wafer were obtained by conventional PR and BARC material on silicon wafer. Exposure pattern for 80nm line and space were curried out with Nikon S307E (Annular, NA=0.85, σ=0.89/0.59). PR: Immersion photo resist-A. BARC: ARC29A (80nm, Nissan Chemical Industries)

3. Measurement of IR Spectra

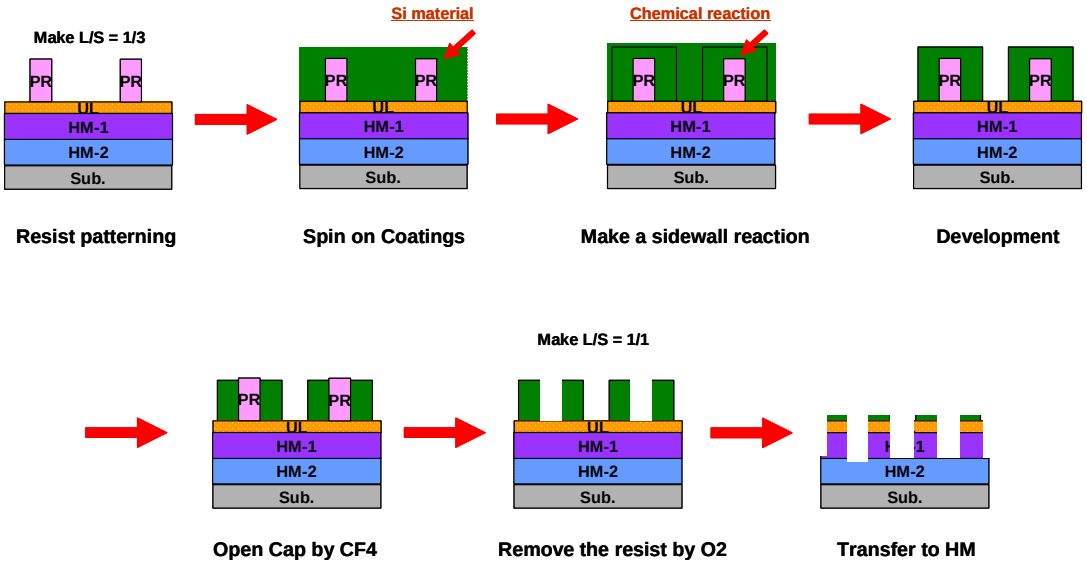
The IR spectra of spin on sidewall material were measured by Nicolet6700（Thermo FISHER SCIENTIFIC) with Variable Angle Reflectance Accessory.

4. Measurement of Interface Roughness, Film Thickness and Film Density

The interface roughness, film thickness and film density of spin on sidewall material were measured by X-ray diffraction（PANalytical). Voltage: 45kV, Current of electricity: 40mA, X-ray wave length: CuK α_1

5. Preparation of Spin on Sidewall

The Process flow of Spin on sidewall is shown as follows in Fig. 1.



From above scheme, requirement of spin on sidewall material are

1. Not mixing with photo resist: Keeping resist profile
2. Make a sidewall reaction by acid of photo resist
   Photo acid will be induce the C.L. reaction
   Need to control profile of sidewall
3. Easy to remove non react area of coatings by some solvent
4. High Si content is needed for Hard Mask

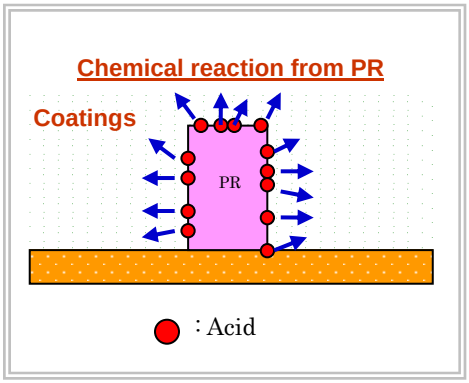


Fig . 1. Spin on sidewall process and concept

Results and Discussion

1. Solvent of Spin on Sidewall Material

After spin on sidewall material coating, photo resist pattern should be keep the pattern size and profile. Therefore, solvent system of BARC (Bottom Antireflective coatings) and photo resist are not acceptable for it. On the other hands, solvent system of TARC (Top Antireflective coatings) and Top coat material are suitable one, we think.

Using top-coat solvents, we checked the solvent resistance with some immersion

resists. These solvents were alcohol type or alcohol derivatives type.

	Top coat solvents		
	solvent-1	solvent-2	solvent-3
Photo resist-A	2	9	10
Photo resist-B	14	69	10
Photo resist-C	48	14	54

unit: Å

Table1. Summary of solvent resistance: the matrix of immersion photo resists and top coat solvents

From this result, we could not found suitable solvent for any photo resist. However, for this investigation, we selected best combination: photo resist-A and solvent-1 and apply below sidewall testing; Test resist pattern was made by photo resist-A, and solvent of spin on sidewall material was applied solvent-1.

2. Investigate of Spin on Sidewall Process

2.1 Thermal Reaction of Si-coatings. As we showed (1.4), spin on sidewall material was coated on resist pattern directly. Then it makes a chemical reaction from resist surface by acid which was remained after lithography. After that, it should be removed by developer. The amount of sidewall reaction was determined to be the difference between the initial resist pattern size and sidewall size after development.

At this test, we tried some samples: common material concept was “Si containing material”, and big material difference was “Molecular weight”.

Test Process of Sidewall material

0. Check the CD size
1. Coat Sidewall material: 1500rpm 60sec
2. Baking: 150℃/60 sec
3. Development: Remove non reaction area
4. Spin dry: 3000rpm 30sec
5. Check the Sidewall CD by CD-SEM: get wider CD if make sidewall

Sample	Mw	CD (nm)	
		initaila	final
1	small	80	80
2	samll	80	80
3	samll	80	83
4	Big	80	81
5	Big	80	79

Table 2. Results of sidewall reaction: material dependency under 150 degree C baking

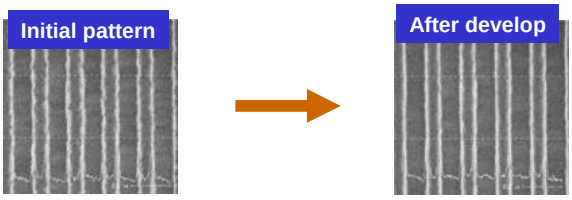


Fig. 2. Top View results of sample 3. Initial pattern size was 80nm. After development, it was not sidewall increasing.

From this result, we couldn’t see CD increasing by sidewall reaction. This phenomenon will be occurred by small amount acid of PR surface. We needed to conceder how to induce and to diffuse acid to coatings for making sidewall.

Ex. 1016, Page 440

2.2 Exposure Irradiation for Inducing Acid. From last results, we think it should be need more acid inducing. So we tried to add exposure irradiation (UV broad band) to original process because it still remains some PAG (photo acid generator) in resist pattern. Using this PAG, we did check the phenomena of sidewall reaction. Detail scheme was shown as below.

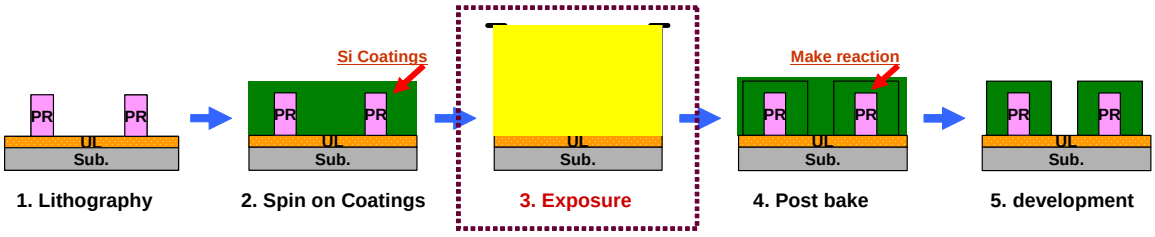


Fig. 3. Spin on sidewall with exposure process

Test Process of Sidewall material

0. Check the CD size
1. Coat Sidewall material: 1500rpm 60sec
2. Exposure for induce the PAG from patterning resist
3. Baking: 150℃/60 sec
4. Development: Remove non reaction area
5. Spin dry: 3000rpm 30sec
6. Check the CD by CD-SEM: get wider CD if make sidewall

Sample	Mw	CD (nm)	
		initaila	final
1	small	80	52
2	samll	80	81
3	samll	80	79
4	Big	80	103
5	Big	80	129

Table 3. Results of sidewall reaction: material dependency under exposure irradiation and 150 degree C baking

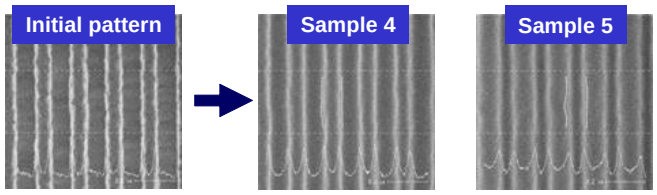


Fig. 4. Top View results of sample 4 and 5.

Especially higher Mw samples (sample 4 and 5) were showed CD increasing (make a sidewall reaction). From these results, we think the amount of sidewall reaction under this condition is not too much. So that, lower Mw compounds is not able to get cross linking networks: difficult to get the wider sidewall pattern. However, higher Mw compounds already have intertwined state in itself and get to enough networks easily, we guess.

3. Development Process after Sidewall Reaction

Regarding a kind of developer, coater friendly process is one of the customer requirements. Testing sample was coated sample 5 sidewall material on 80nm resist line pattern using photo resist-A. After UV exposure and baking (150degree C for 60 sec), we check the developer dependency.

Ex. 1016, Page 441

In this section, we checked the below three solvents:

1) Solvent-1 (top coat solvent): Solvent of Sidewall materials
2) PGMEA: Resist solvent
3) Conventional developer + DIW.

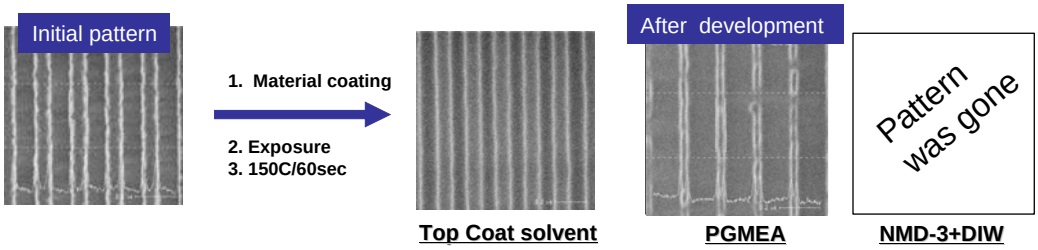


Fig. 5. Top View results of developer dependency.

In case of solvent-1 (top coat solvent), it could keep the sidewall shape. On the other hands, PGME and NMD-3 + DIW case, Sidewall pattern was shrieked or gone. We don't understand the reason why only MIBC can be suitable for developer until now.

4. Study of Controlling Sidewall CD

Under exposure irradiation and baking process, sidewall reaction can be obtained. In order to control sidewall width more severely, we studied the dependency of process condition: (a) Exposure irradiation (ArF, UV), (b) Post exposure baking. Testing sample was coated Si-sample 5 sidewall material on 80nm line pattern of photo resist-A.

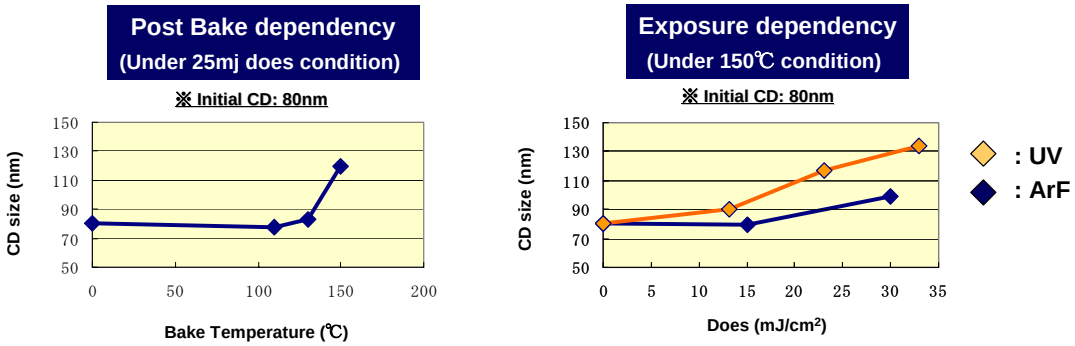


Table 4. Results of sidewall reaction: Process dependency of exposure irradiation and baking temperature

Regarding Post baking temperature, Over 130 degree C is effective for making sidewall. Generally, acid diffusion in the resist is related with the bake temperature and time. This result also showed same trend. And as for Exposure irradiation, UV irradiation has bigger CD increasing than ArF one. One of the reasons is absorption of Si-materials. In Si-material 5 case, it has alos0 some absorption at 193nm. For make a clear it, we need to check the different absorption sample using same chemistry.

Below figure was showed cross section result after development. This resist pattern was used to nearly 1 by 3 line and space. As you can see, sidewall material covered on

Ex. 1016, Page 442

resist pattern. However, Si-sidewall layer was showed different thickness: it means thickness of number 4 and 5 showed not same thickness (about 5 nm differences). This one will be cause the etching issue for make a Si-sidewall pattern 1 by 1 line and space. In future, we would like to study controlling sidewall reaction from process and chemical side.

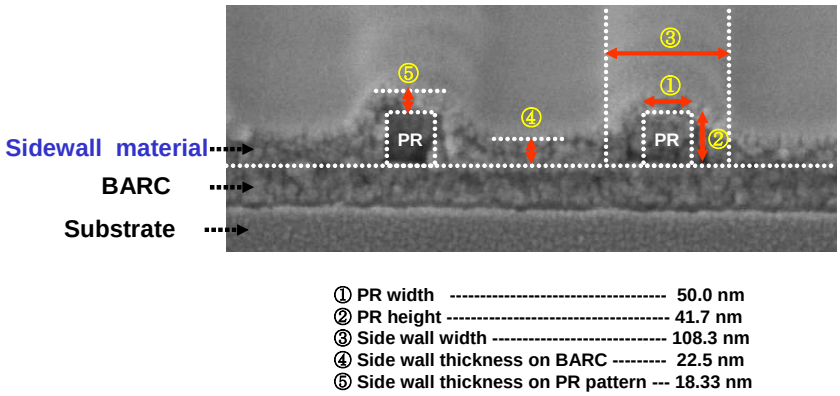


Fig. 6. Cross section View of sidewall after development

5. Research in Sidewall Reaction

5.1 Sidewall Reaction by IR Spectra. For make a clear the sidewall reaction, we investigated it by IR spectra. And for basic study, we prepared model sample: Si-compound adding catalyst and chased the IR spectra under baking from 100 to 280 degree C. There were some peaks changing: the decreasing of Si-OH peak about 900 cm^{-1} and the increasing of Si-O-Si peak about 800 cm^{-1} (Fig. 7. left figure).

Moreover, more close to real stuck, we tried to check it using model stuck: Si-sample 5 on silicon wafer coated organic film including PAG as below. After UV irradiation, we did chase the IR spectra under baking from 70 to 150 degree C. the spectra difference was very small (Fig. 7. right figure). It is a significant difference or not, we could not make sure now.

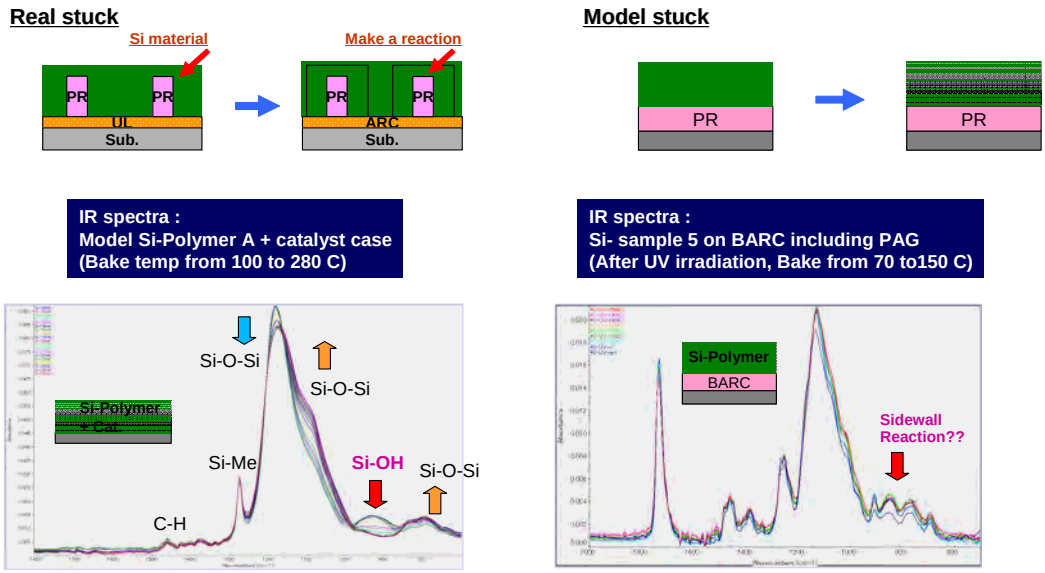


Fig. 7. Left IR spectra was shown the reaction of model film: Si-material and catalyst. Right IR spectra was shown the reaction of model stuck: Si-sample 5 on silicon wafer coated organic film including PAG.

5.2 Sidewall Reaction by X-ray Diffraction. Firstly, we prepared the below stuck for check (a) the inter face between PR and Si-sidewall material, and (b) film density of Si-material after sidewall reaction by X-ray diffraction.

Test sample stuck: Si-sample-5 on silicon wafer coated photo resist-A. Sample treatment for test (b): UV irradiation 30mJ/cm^2, baking temperature150 degree C for 60sec

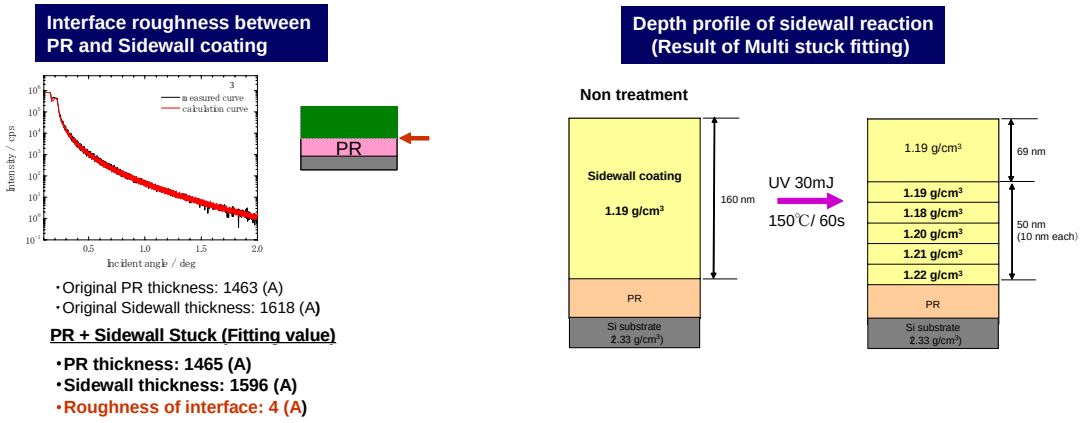


Fig. 8. Spin on sidewall process and concept

About interface roughness between photo resist-A and Si-material 5, there were only 4 A roughness. This was very similar result of **3.1.** But, as for film density of Si-sidewall material, they had very few depth profiles. It is necessary to study with different formulation and process condition to discuss detail.

Conclusions

We investigated spin on type sidewall coatings. Higher Mw Si polymer is suitable for making sidewall under this process. And top coat solvent can be removed the non reaction area for keeping sidewall shape.

Regarding process condition, exposure energy and bake temperature are one of the factor of controlling Sidewall CD. And we also studied sidewall reaction by IR spectra and X-ray diffraction. From this study, we got some basic date but it still needs to keep do it for controlling sidewall reaction.

Acknowledgments

The authors would like to express their thanks to Brewer Science, Inc

References

1. M. Khojasteh, et al., “Building an immersion topcoat from the ground up: materials perspective” *Proc. SPIE* **6519**, 651907 (2007)
2. M. Hori, et al., “Sub-40nm Half-Pitch Double Patterning with Resist Freezing Process” *Proc. SPIE* **6923**, 69230H (2008)
3. E. Nishimura, et al., “Precise CD control techniques for double patterning and sidewall transfer” *Proc. SPIE* **6924**, 692425 (2008)

ECS Transactions, 18 (1) 419-425 (2009)
10.1149/1.3096480 © The Electrochemical Society

Residue Testing of Developer-Soluble Bottom Anti-Reflective Coatings

Carlton Washburn, April Evers, and Brian Smith

Brewer Science, Inc., Rolla, MO 65401, USA
cwashburn@brewerscience.com

Organic developer-soluble bottom anti-reflective coatings (DBARCs) are commonly used in the photolithographic process for implant and high-k/metal gate integration. They provide added reflectivity control over topography compared top anti-reflective coatings (TARCs). DBARCs are coated onto substrates and then removed during the develop step with the resist. Occasionally a thin post-develop residue (PDR) remains after the DBARC removal process. The main drivers of this residue are DBARC bake temperature, followed by develop time. For this investigation the amount of residue was tested using Si, SiO_2, native SiO_2, SiN, and HfO_2 substrates at different bake temperatures. In addition, contact and proximity baking are compared and shown to be equivalent in terms of their effect on PDR. This work demonstrates that optimizing the bake process of a DBARC is an effective way to control PDR on a variety of substrates.

1.0 Introduction

As integrated circuit (IC) technology advances, the materials used in manufacturing become more sophisticated. As materials advance, the processing required to optimize and qualify the materials becomes more complex. The photolithography process engineer in particular faces many challenges while selecting and using resists and anti-reflective coatings. This complexity places more pressure on the process engineer to develop processes that are both robust and cost-effective. There are now several different types of resists, hardmasks, gap-fill materials, underlayers, and bottom anti-reflective coatings (BARCs). BARC varieties include traditional dry-etch, wet-developable, graded index, and silicon-containing (1,2,3,4).

Process development for developer-soluble bottom anti-reflective coating (DBARC) materials is distinctly different from that for traditional dry-etch BARCs (5). Dry BARC process development is straightforward and has fewer variables. Once resist compatibility is established, the remaining issues involve mainly thickness, etch selectivity, and defect concerns (6). DBARC process development must take into account many additional concerns such as the choice of thermosetting or photosensitive materials (5), as well as the control of post-develop residue (PDR), which can vary depending on the substrate, and is the focus of this investigation.

Two typical applications for DBARC are in ion implantation and high-k/metal gate (HKMG) patterning. These applications can be negatively affected by PDR unless methods exist to characterize the process at an early stage. Understanding the causes of residue and its impact on processing can help to reduce integration problems.

Ex. 1016, Page 447

Understanding what steps to monitor becomes critical to maintaining a stable process. This paper will show that by use of process optimization, PDR with thermosetting DBARC can be minimized. It will further evaluate how this can be applied to different substrates common to implant and HKMG applications.

2.0 Experimental Method

The method used to evaluate PDR was based on spectroscopic ellipsometry (SE). Using an M-2000® SE from J.A. Woollam Co., the process begins with a measurement of the native oxide or inorganic film thickness on the substrate. Next, the DBARC, Brewer Science™ ARC® DS-K101 coating, was dispensed onto the wafer. Cauchy coefficients of the DBARC were then recorded from the SE and the wafer was developed in 0.26N TMAH with a surfactant (7). Finally, the substrate was measured again with the SE. These measurements were used to build a model that looks for subtle changes in the polarization of the light that would be caused by a residue. If a residue was found, then the average thickness was calculated.

SiO_2, native SiO_2, SiN, and HfO_2 substrates were used for this study because they are representative of some of the surfaces encountered in implant and high-k/metal gate lithography. DBARC was dispensed onto 200-mm wafers by hand using a TEL Mark 8 Clean Track coater system with HHP proximity bake plates. The develop process used an E^2 developer nozzle. Also, 100-mm wafers were hand-coated and processed using a Brewer Science™ Cee® 200CB system and a contact bake plate. Typical bake conditions were 185°C for 60 seconds, followed by developer-rinse for 60 seconds, and final deionized water rinse.

Factors that can contribute to residue were screened and modeled with a design-of-experiments (DOE) approach. Stat-Ease's Design Expert® version 7.0.1 software was used for DOE design and analysis. DOE is a statistical tool that allows screening of effects from multiple factors more efficiently then by testing one factor at a time. Fractional factorial DOE designs offer a straightforward approach for evaluating main effects and some two-way interactions, while follow-up response-surface DOEs provide detailed information on the best locations to optimize a process.

3.0 Results and Discussion

Using the SE method, six process factors and one chemical loading factor that adjusts develop rate were evaluated as possible residue drivers. A resolution-IV fractional factorial DOE was used to screen for the drivers which had the greatest impact on PDR. The design is detailed in Table 1. Because the goal was to screen a large number of factors 100-mm substrates were used initially for speed and convenience.

TABLE I. Factors utilized in the DOE.

Factor	Unit	High	Low
Bake Temp	(°C/60 s)	205	135
Bake Time	(s)	105	15
Develop Time	(s)	120	30
Film Thickness	(nm)	290	750
Exposure	(mJ/cm^2)	60	0.0
Develop Puddles	(#)	1	4
Develop Rate	(ratio)	High ratio	Low ratio

The resulting PDR data were processed using Design Expert®. The half-normal plot is shown in Figure 1. Bake temperature had the greatest effect, with bake time and develop time having lesser effects. These three factors were used to design a central composite response-surface DOE to map out the factor interactions and to generate and study trends.

The response-surface DOE was run twice, once on 100-mm contact-baked substrates and again using 200-mm proximity-baked wafers. Because bake temperature was determined to have the greatest impact on PDR, the goal was to compare the resulting response surfaces to determine whether or not there was a difference in trends from the two different heat transfer modes. A second goal was to gauge any processing offset between contact and proximity-baked wafers. Results of the two DOEs are shown in Figure 2. The curvature of the surfaces shows that the trend for both experiments was the same. Higher bake temperatures increase PDR. Also, develop time had an interaction with bake temperature. At very high bake temperatures and short develop times, the remaining BARC film increased rapidly. This result was not so much a function of residue, rather incomplete development of a highly crosslinked film. Chemically the degree of crosslinking increases with an increase in bake temperature, but not necessarily bake time, as with Arrhenius behavior. This is confirmed by the relatively insignificant effect of bake time, which is not shown.

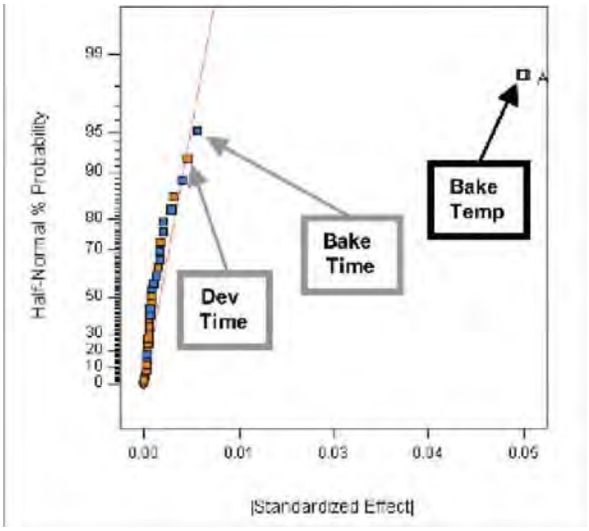
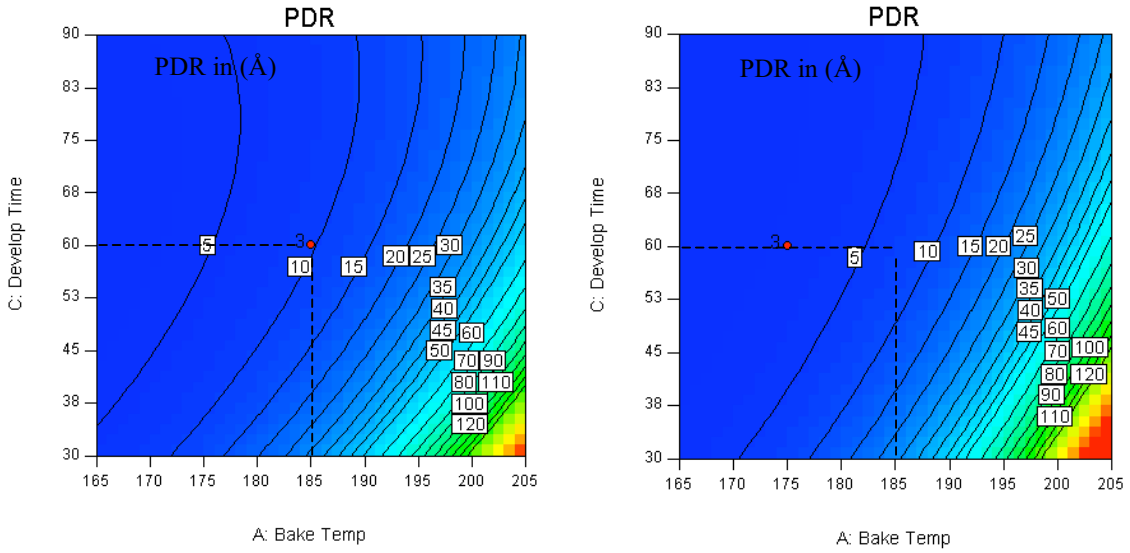


FIGURE 1. Factorial DOE half-normal plot.

A slight offset occurred between the center of the two surfaces, which is shown by the dashed lines in Figure 2. The standard deviation of the replicated center points for the 100-mm DOE was 0.6 Å, and for the 200-mm track process DOE the standard deviation was 0.3 Å. Thus the offset between the two response surfaces is beyond the experimental error and is real. This offset likely came from a difference in temperatures between the two hot plates used. Because different bake temperatures were used during the testing, hot plate equilibrium times may not have been sufficient. This is the likely source of offset between the two tests.



200-mm proximity baked 100-mm contact baked

FIGURE 2. DOE response surface plots of PDR.

To confirm the 200-mm DOE results, additional wafers were processed. The develop time and bake time were both held constant at 60 seconds. Then the resulting PDR data were plotted against the DOE model to confirm the model (Figure 3). The model and empirical data match well up to 180°C, at which point they separate, then recombine near 205°C. This mild deviation is expected from comparing the model fit against the new, more complete data set, and this confirmation was considered successful. It also showed that optimizing DBARC bake temperature can minimize PDR. This will improve HKMG and implant applications that are sensitive to residue.

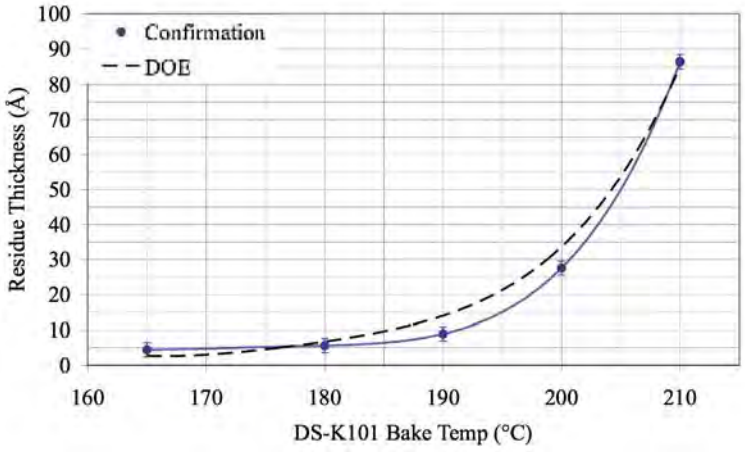


FIGURE 3. Bake confirmation of 200-mm track-processed wafers.

Previous testing has shown that photosensitive DBARCs have different amounts of PDR when processed over silicon compared to silicon dioxide (7). To further the understanding for the thermosetting DBARC (ARC® DS-K101 coating), similar testing was performed. Silicon and 200 nm of SiO_2 were tested to compare to previous photosensitive DBARC data. 300 nm of SiN and 7 nm of HfO_2 were also tested. The results are shown in Figure 4. By far, SiN showed the most residue. SiN is well known interact with organic resist layers to cause scumming problems due to the residual N-H bonds that neutralize the acid in chemically-amplified resists (8). A similar mechanism may be at work in the case of DBARC, resulting in a neutralizing of the base-soluble components at the DBARC/nitride interface. Residue thickness from HfO_2 was half that of silicon, with SiO_2 showing the least amount of residue. The reduced residue result with SiO_2 aligns with previous results from photosensitive DBARC testing (7).

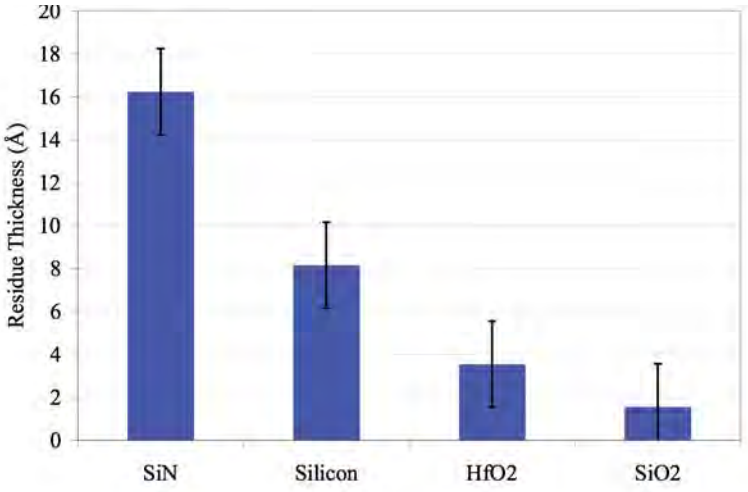


FIGURE 4. PDR on different substrates.

To further explore the relationship between SiO_2 substrates and residue, the native oxide of a silicon wafer was mapped, then ARC® DS-K101 coating was applied to the wafer and developed away. These results are shown in Figure 5. A linear trend resulted, indicating that increases in native SiO_2 produce less PDR. The proposal that a thick film of SiO_2 would decrease the surface silicon sites available for DBARC to bind to was previously proposed (7). This concept was supported by these results in that as the native oxide increases, it may become less and less ordered like pure silicon.

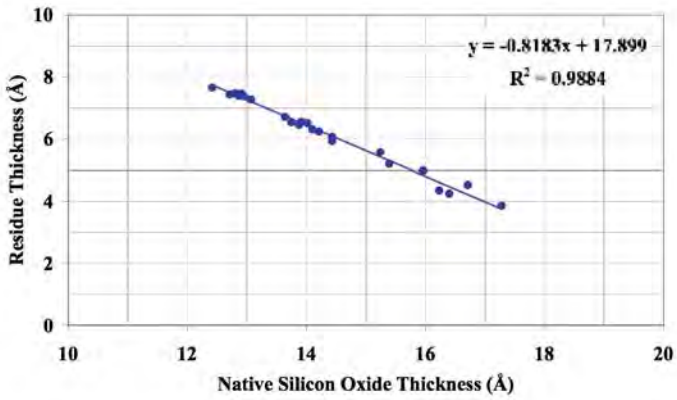


FIGURE 5. PDR on various thickness of native SiO_2.

4.0 Conclusions

To summarize, a factional factorial DOE was designed to screen several processing and one chemical factor for residue using ARC® DS-K101 coating, a thermosetting DBARC. It was learned that bake temperature was the main driver. Two response-surface DOEs were then designed to further investigate baking as a residue driver. Bake temperature was confirmed to be the largest driver, followed by an interaction with develop time. This showed that optimizing DBARC process conditions can minimize PDR for applications including HKMG and implant patterning.

Substrate testing showed SiN to have the largest effect on residue, followed by Si, HfO_2, and SiO_2, which showed the least residue. Native SiO_2 was tested and showed a linear trend of decreasing PDR with an increase in native oxide.

5.0 Acknowledgments

The authors would like to thank Brandy Carr for her wafer processing assistance. The authors would also like to thank Andrea Fuchser of J.A. Woollam Co. for her advice and support during this study.

6.0 References

1. Xie Shao, Alice Guerrero, and Yiming Gu, "Wet-Developable Organic Anti-Reflective Coatings for Implant Layer Applications," SEMICON China 2004 SEMI Technology Symposium, March 17-19, 2004.

2. S. Lee, J. Jung, S. Cho, C.Lim, C. Bok, H. Kim, S. Moon, J. Kim, *Proceedings of SPIE*, Vol. 6153, 2006.
3. S. Burns, D. Pfeiffer, A. Mahorowala, et al, *Proceedings of SPIE*, Vol. 6153, 2006.
4. D. Goldfarb, S. Burns, L. Vyklicky, et al, *Proceedings of SPIE*, Vol. 6923, 2008.
5. C. Washburn, A. Guerrero, R. Mercado, D. Guerrero, J. Meador, *INTERFACE 2006: Proceedings of the 43rd Microlithography Symposium*, 2006.
6. A. Gabor, S. Halle, C. Kallingal, *Proceedings of SPIE*, Vol. 5753, 2005, pp. 699-707.
7. D. Guerrero, J. Lowes, R. Mercado, *Journal of Photopolymer Science and Technology*, Vol. 20, no. 3, 2007, pp. 339-343.
8. J. Chun, C. Bok, K. Baik, *Journal of Photopolymer Science and Technology*, Vol.10, no.3, 1997, pp.445-450.

CHAPTER 5

CMP & CLEANING

ECS Transactions, 18 (1) 429-434 (2009)
10.1149/1.3096481 © The Electrochemical Society

Investigation on Chemical Mechanical Polishing of GeSbTe for High Density Phase Change Memory

Min Zhong [a,b], Zhi-Tang Song [a], Bo Liu[a], Song-Lin Feng [a], Zhi-Feng Xie[c], Fu-Xiong Zhang [c], Yang-Hui Xiang[c]

[a] State Key Laboratory of Functional Materials for Informatics, Laboratory of Nanotechnology, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050
[b] Graduate School of the Chinese Academy of Sciences, Beijing 100049
[c] Semiconductor Manufacturing International Corporation, Shanghai 201203

Phase change memory (PCM) is considered to be a prospective candidate to replace flash for next generation non-volatile memory. In this paper, we studied GST chemical mechanical polishing (CMP) process characterization and optimization on 8-inch wafers. CMP experiment was performed on GST blanket and pattern wafers. GST CMP process optimization experiment was made in both chemical (slurry) and mechanical (pressure) aspects. AFM data shows that the RMS of roughness has been reduced from 2.151 nm to 1.276 nm after GST CMP on the blanket wafer. On the other hand, high density (40%) GST pattern wafer was polished to evaluate the CMP capability. SEM and EDS results demonstrate that GST via array has been achieved by GST CMP process while the element stoichiometric proportion in GST via is not changed. This work paved the way for PCM development to the commercial high density PCM devices in the future.

Introduction

Phase change memory (PCM) is one of the most prospective candidates to replace flash for next generation non-volatile memory due to its outstanding scaling, high speed, great endurance characteristics and compatibility with standard silicon processing [(1,2)]. In these years, many famous semiconductor companies such as Samsung, Ovonyx, and Intel have made remarkable progress in PCM technology [(3)]. PCM bases on the reversible switching of phase change film between amorphous (reset-high resistance) and crystalline (set-low resistance) states to enable the data storage. $Ge_2Sb_2Te_5$ (GST) alloy is the most widely used phase change material for PCM application. Recently, a confined structure is thought to reduce the reset current and thermal cross talk with a simple integration process comparing with the conventional planar structure (i.e., T-shape) [(4)]. Therefore, Chemical mechanical polishing (CMP) process of GST alloy is the critical process step to make a confined structure and obtain the global planarization of the whole wafer. Our group reported

Ex. 1016, Page 457

GST CMP with our self-made slurry [5]. However, no detailed information about GST CMP on 8-inch wafers has been presented until now.

In this paper, we investigated GST CMP process characterization and optimization on 8-inch wafers. CMP experiment was made on GST blanket and pattern wafers. GST CMP process optimization experiment was carried out in both chemical (slurry) and mechanical (pressure) aspects. The roughness of the surface before and after CMP was measured by atomic forced microscopy (AFM) scanning. On the other hand, high density (40%) GST pattern wafer was polished in the optimized condition to evaluate the capability of GST CMP process. Top view and cross-sectional images by scanning electron microscopy (SEM) and element composition measurement of the GST via by energy dispersive X-ray fluorescence spectroscopy (EDS) was performed to verify GST CMP performance.

Experimental

8-inch oxide wafers deposited about 460 nm GST film were polished using a CMP tool (EPO222, Ebara) with a soft pad (Politex, Rohmhaas) and acidic silica slurry. Considering the previous research result on GST CMP process [6], some additive was added in slurry. The polishing parameter of the optimized CMP process was summarized in Table 1. The polishing rate was obtained through GST film thickness before and after CMP. The roughness of the polished surface was measured by AFM (TM250, Q-Scope).

High-density GST pattern wafer was applied to evaluate the capability of GST CMP process. GST thin film was deposited on the pattern wafer by RF sputtering. GST via critical dimension (CD) is 600 nm and via space CD is 400 nm. Initial GST via depth is 150 nm. Then, the wafer was polished in the optimized condition. SEM and EDS (S-4700, Hitachi and INCA Energy, Oxford) were used to check the surface of the polished wafer and test the element composition of GST via.

Table Ⅰ. The polishing parameter of CMP

Parameter	Index
Polishing pressure (psi)	2~4
Velocity of polishing table (rpm)	60
Velocity of wafer (rpm)	57
Flow rate of slurry (ml/min)	300~500

Results and Discussions

Blanket Wafer Polishing

The CMP process is a combination of chemical reaction and mechanical abrasion. In order to attain a planarized, damage-free surface, the balance of chemical reaction and mechanical effect is necessary. In chemical reaction aspect, GST oxidization plays an important role in GST CMP process. GST material removal rate (MRR) is very sensitive to the concentration of the oxidant. Details of the oxidant effect on GST

Ex. 1016, Page 458

CMP performance and its mechanism have been presented previously [6]. With the particular oxidant addition slurry, we adjusted the polishing parameters such as pressure, rotation rate and slurry flow rate (shown in Table 1). The range of the pressure is from 2 psi to 4 psi, while the polishing table rotation rate and the wafer rotation rate is 60 rpm and 57 rpm, separately. Although high pressure and rotation rate can increase GST MRR, the moderate pressure and rotation rate is chosen to get the best quality polished surface. Likewise, slurry flow rate is tuned from 300 ml/min to 500 ml/min to gain the good polishing performance.

TABLE Ⅱ. The pre-CMP and post-CMP thickness of GST film and the removal rate of GST in different position of 8-inch wafer

Item	Center	Left	Right	Top	Bottom	Average	Range
Pre Thk (A)	4673	4652	4812	4560	4589	4657	252
Post Thk (A)	2532	2665	2551	2283	2641	2534	382
MRR(A/min)	2141	1978	2261	2277	1948	2123	329

According to GST material removal rate (MRR) and the roughness of the polished surface, an optimized condition was confirmed. The thickness of pre-CMP and post-CMP and MRR of GST in 5 positions of the wafer are listed in Table 2, separately. MRR of GST is 212 nm and its range is 33 nm. The uniformity of GST MRR is acceptable to an 8-inch wafer. 3-D image of surface for pre-CMP and post-CMP over 100 μm^2 by AFM scanning are also shown in Figure 1 (a) and (b). The initial surface roughness (RMS) is about 2.151 nm, and the RMS of the post-CMP surface is reduced to 1.276 nm. The AFM results indicate that GST film can be removed effectively leaving a smooth and damage-free surface.

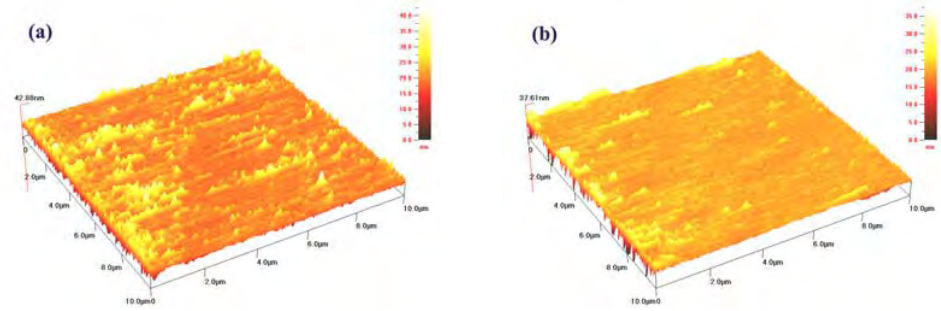


Figure 1. 3D image of the surface (a) pre-CMP (b) post-CMP

Pattern Wafer Polishing

For pattern wafer, GST polishing process is more complex. A certain amount of over polishing during GST CMP is required to make sure all the GST clear on the surface. During the over polishing, both GST and SiO_2 are present on the surface and are polished simultaneously, but with the different polish rates. Dishing and erosion create a non-planar surface which could greatly effect on the consistency of PCM

Ex. 1016, Page 459

cell's electrical characters [7]. So a good multi-step CMP process control is critical to avoid or reduce GST dishing and oxide erosion.

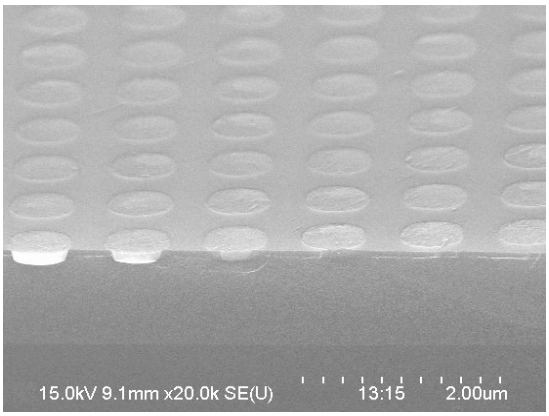


Figure 2.The surface topography of the polished GST pattern wafer

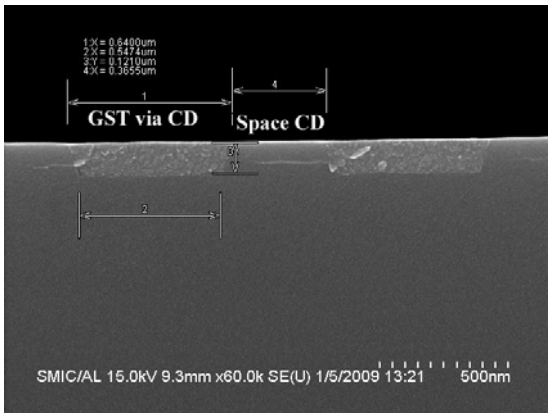


Figure 3. Cross-sectional SEM image of GST pattern wafer

A high density GST pattern wafer was polished with the fine tuning condition. Figure 2 shows the surface topography of the polished wafer. No GST residue or serious scratch was found on the surface. The image shows that high density of GST via array has achieved by GST CMP process. As shown in Figure 3, GST via CD was 640nm and via space CD was 365 nm. The pattern area density is about 40%, which is high density in PCM device application. Also, the polished GST via depth is 120nm without obvious dishing. The oxide loss is about 30 nm, which is acceptable in our present PCM process. The element analysis of the GST via shown in Figure 4 was carried out by EDS. The atomic percentage of element Ge:Sb:Te is 24:23:53 and the material is considered as $Ge_2Sb_2Te_5$ (table 3). The EDS result shows that GST stoichiometric proportion is not changed after CMP process. Above all, high density confined structure of GST cell array has been achieved by GST CMP process. In addition, oxide erosion which is strongly pattern dependent is studied in another paper.

Ex. 1016, Page 460

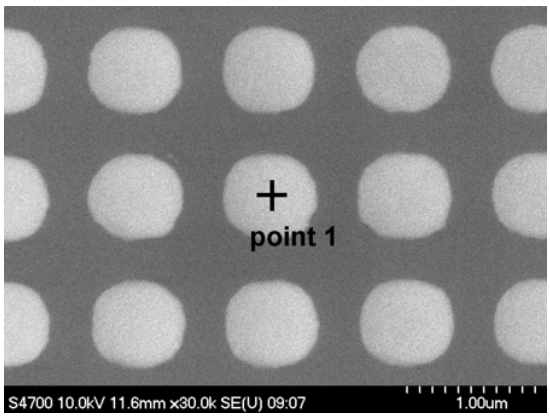


Figure 4.Top view SEM image of the polished GST pattern wafer

TABLE Ⅲ. Element composition of GST in the oxide via by EDS

Element	Mass percentage (%)	Atom percentage (%)
GeK	15.69	24.39
Sb L	24.89	23.07
Te L	59.42	52.54
Total	100.00	100.00

Conclusions

In summary, through the optimization of GST CMP process, the RMS roughness of the polished surface was reduced from 2.151 nm to 1.276 nm while GST removal rate is 212nm/min on the blanket wafer. For pattern wafer, SEM and EDS results demonstrate that high density (40%) GST via array has been achieved by GST CMP process while the element stoichiometric proportion in GST via is not changed. The GST CMP process capability is also discussed. This work paved the way for PCM development to the commercial high density PCM devices in the future.

Acknowledgments

The work is supported by National Basic Research Program of China (2007CB935400 and 2006CB302700), National High Technology Development Program of China (2008AA031402, 2006AA03Z360), Science and Technology Council of Shanghai (0652nm003, 0752nm013, 07QA14065, 07SA08), and National Nature Science Foundation of China (60776058).

References

1. Ovshinsky SR., *Phys. Rev. Lett.*, **21,** 1450 (1968)
2. Lai S., *IEDM Tech. Dig.* , 255 (2003)
3. J.H. Oh, J.H. Park, Y.S. Lim, H.S. Lim, Y.T. Oh, J.S. Kim, J.M. Shin, Y.J. Song, K.C. Ryoo, D.W. Lim, S.S. Park, J.I. Kim, J.H. Kim, J. Yu, F. Yeung,

Ex. 1016, Page 461

C.W. Jeong, J.H. Kong, D.H. Kang, G.H. Koh, G.T. Jeong, H.S. Jeong, Kim Kinam, *IEDM*, 1-4 (2006)

4. Jin-Hyung Park, Jong-Young Cho, Hee-Sub Hwang, Ungyu Paik, Jea-Gun Parka, *Electrochemical and Solid-State Letters*, **11** (10), H288-H291 (2008)
5. Qi-Bin Liu, Zhi-Tang Song, Kai-Liang Zhang, Song-Lin Feng, Bomy Chen, *Chinese Physics Letters*, **23**, 8, 2296-2298 (2006)
6. M. Zhong, Z.T Song, B. Liu, L.Y. Wang, S.L. Feng, *J. Electrochem. Soc.*, **155** (11), H929-H931 (2008)
7. M. Zhong, Z.T Song, B. Liu, S.L. Feng, B. Chen, *Electronics letters.*, **44** (4), 322-323 (2008)

ECS Transactions, 18 (1) 435-440 (2009)
10.1149/1.3096482 © The Electrochemical Society

Study on CMP Mechanism and Technology of Sapphire Substrate for Photo-conducting Device

X.H. Niu[a], T. Wu, X. H. Zhao, B. M. Tan, Y. L. Liu

Institute of Microelectronics, Hebei University of Technology, Tianjin, 300130, China

Sapphire(α-Al_2O_3) single crystals are widely used in many areas because of the special properties. As an important photoconducting device substrate material, stringent surface quality requirements are required. The use of CMP technique can produce high quality surface finishes at low cost and with fast material removal rates. In this paper, the polishing mechanism was studied. According to Arrhenius equation, temperature is the important factor of influencing chemical reaction rate. So the chosen temperature was 45℃. SiO_2 sol was chosen as abrasive, and the particle size is 40nm. For ensuring the stability of abrasive and chemical action of polishing slurry, the pH value was determined at 11.5~12. After polishing and cleaning the sapphire surface, the measured removal rate was above 11 μ m /h and the surface roughness was Ra 0.3nm. From the results, it was found that using such method, the optimal sapphire surface can be gotten.

Introduction

Sapphire, which is a single crystal form of a-alumina (corundum), is an important ceramic material used widely in a range of applications such as optics, electronics, and temperature sensing. In many of these applications, stringent surface quality requirements (i.e., finish and flatness) are required. Low subsurface damage is also critical. It has been established that the crystal structure of epitaxial films is strongly influenced not only by the substrate material and its orientation, but in a great extent also by the surface preparation of the substrate [1-4]. Fine surface machining and polishing of ceramics for opto-electronic applications may exceed 80% of total cost. The use of chemically assisted polishing techniques such as chemical mechanical polishing (CMP) may produce high quality surface finishes at low cost and with fast material removal rates, little or no surface and subsurface damage, elimination of surface defects, high efficiency[5-8].
However, many ceramics such as sapphire are noted for wear and corrosion resistance due to the high chemical and thermal stability, high hardness and in many cases the effective CMP processes are not available or not understood [5-7, 9-11]. So the influence of slurry and process parameters on sapphire substrate removal rate and surface quality become the main research subject. In this paper, the CMP mechanism of sapphire substrate was studied. The influence of pH value and different temperature on removal rate was discussed. And the optimal polishing process was proposed.

CMP Mechanism of Sapphire Substrate

CMP is related to mechanics, hydrodynamics, physical chemistry, material tribology and other subjects, its kinetics process and control process is still not very clear. The most basic polishing rate equation that has been widely used is the so-called Preston's

Ex. 1016, Page 463

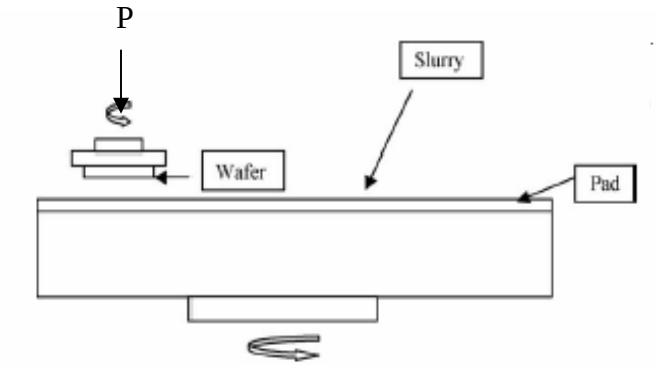


Fig.1 Schematic illustration of a CMP system

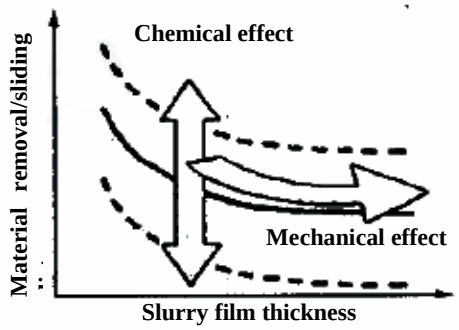


Fig.2 Mechanical and chemical effects on the material removal in CMP

Equation [12], although there are various models for different aspects of a CMP process. The Preston equation which predicts that the removal rate ($RR=\triangle h/\triangle t = KpPV$) , i.e., the thickness decrease over time ($\triangle h=\triangle t$/)[12], depends linearly on the downward wafer pressure P and the relative velocity between the pad and the wafer surface V, where Kp is Preston's coefficient, which is a strong function of the other CMP parameters. The wide use of Preston's equation for CMP is surprising since it was obtained for polishing with hard pads, and emphasized mechanical action for CMP process. But chemical action is more complex and has larger influence on CMP process. It was found in our investigation that removal rate in CMP process was not the simply adding of mechanical and chemical action, it was $R >> R_M + R_C$.

Chemical mechanical polishing(CMP) is the combination technique of mechanical grinding and chemical corrosion. During the polishing process, it includes chemical reaction kinetics process and mass transmission process. Fig1 shows the schematic illustration of a CMP system. Fig2 shows the mechanical and chemical effects on material removal during CMP.

Chemical Kinetics Process

During the sapphire substrate CMP process, under the alkali media the chemical reactions as follows[5]:

$$Al_2O_3 + H_2O \Leftrightarrow 2AlO(OH) \quad (1) \qquad Al_2O_3 + 3H_2O \Leftrightarrow 2Al(OH)_3 \quad (2)$$

$$Al_2O_3 + 2OH^- \Leftrightarrow 2AlO_2^- + H_2O \quad (3) \qquad Al(OH)_3 + OH^- \Leftrightarrow AlO_2^- + 2H_2O \quad (4)$$

First α- alumina reacts with H_2O and forms the hydrates of α-alumina[5], and the hardness of such hydrates is lower than α-alumina. At the same time, a part of substrate material and hydrates react with alkali, and form water soluble AlO_2^-. Alumina is an amphoteric oxide, so it can react with acid, also react with alkali. In this study, we chose alkali media, the reasons as follows: First alkali has passivating action, so it can avoid etching equipment; Second, if we choose acid media, there will be the reactions as follows:

$$Al_2O_3 + 6H^+ \Leftrightarrow 2Al^{3+} + 3H_2O \quad (5) \qquad Al(OH)_3 + 3H^+ \Leftrightarrow Al^{3+} + 3H_2O \quad (6)$$

$$AlO(OH) + 3H^+ \Leftrightarrow Al^{3+} + 2H_2O \quad (7)$$

From the above chemical formula we can see that alum ion can be formed, which causes ionic secondary pollution. It takes more difficult for cleaning; Third, alkali media takes important action for abrasive stability.

Mass Transmission Process

In such process, there appears grinding between SiO_2 abrasive and polished surface, and the resultants on polished surface are taken away from the machined surface. Under the action of the flowing slurry and running pad, the resultants can be taken away from the platen.

Experimental procedures

The sapphire crystals with c (0001) orientation were obtained by Czochralski (CZ) method. The CMP experimental was performed on C6382I-W/YJ single side polisher (Fig.1). We chose alkali slurry and took solubility nano-SiO_2 sol as abrasive. The particle size was 40nm, which was measured by Marlven Zetasizer HS3000 particle size analyzer, and the concentration was 20wt%. The pressure were 0.05MPa and 0.16MPa, rotate speed was 60rpm and polishing time was 60mins. The RMS roughness was characterized with atomic force microscopy (AFM).

Results and Discussions

For sapphire CMP, the key factors of influence removal rate and surface quality are slurry and process, such as abrasive, medium, regulators, pressure, temperature, pH, and so on. Some factors had been discussed in the before papers[13]. Here the influence of temperature and pH on CMP quality were discussed.

Selection of Temperature

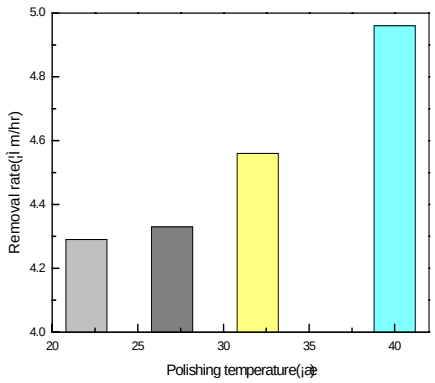


Fig.3 The removal rates with polishing temperature changing

Table1 Variation of temperature with polishing time

		T（℃）			
		A	B	C	D
Time	0	33.8	35	35.6	42
	5min	37.6	38.4	40	45
	10min	39.9	40.7	43	
	15min	42	43	45	
	20min	45	45.2		
Removal rate(μm/h)		8.3	8.99	11.24	12.53

Chemical mechanical polishing is a complex reaction process. The key of determining the optimal process parameters is to achieve high removal rate and good surface quality. Figure3 shows the changing tendency of removal rate with polishing temperature increasing, and Table1shows the tendency of removal rate with different polishing time to the same polishing temperature. From the Figure3 and Table1, it can be seen that the removal rate increases with the temperature increasing and the time to the certain temperature shorting. When the polishing temperature is above 30 , the removal rate increases rapidly. The reason is that under the same pressure and relative velocity action, temperature is the key factor to influence chemical action. That is to say, under the same mechanical action condition chemical action is the slower process, which is the key factor of controlling kinetics process. Arrhenius equation

$$V = K \exp(-\frac{E_a}{RT}) \quad (1)$$

can reflect the rate of chemical action, Activation energy Ea is the difference between the lowest energy of activated molecules and the average energy of molecules. That is to say, the lowest energy of activated molecule is lower, activation energy Ea is lower. During CMP process, with the function of pressure and rotate, reactant molecule can be activated at lower energy with temperature increasing, which can lead to the total activated molecule number increasing and lead to Ea decreasing. According to Arrhenius equation, the rate of chemical action V will be accelerated clearly. Then the chemical and mechanical action can be matched effectively, and the removal rate can be increased.
It also shows the shorter the time to the certain polishing temperature, the higher the removal rate. So during the practical manufacture, the starting time of temperature should be shorten.

Selection of pH value

PH value is an important factor during CMP process. It takes important action on chemical reaction of CMP process, and it must ensure the stability of silica abrasive. As usual, higher pH value is advantage to chemical reaction (shown as fig4). With the pH value rising, the stability of silica abrasive will be lower. When the pH value reaches a certain value, silica abrasive will begin to aggregate. The aggregates can cause polishing pad jamming, which influences the slurry transmitting. Then it causes chemical reaction slower, removal rate decreasing and surface quality lower. According to our experiment results, the pH value range is 11.5～12, which can ensure the floatability of silica abrasive in slurry and high removal rate.

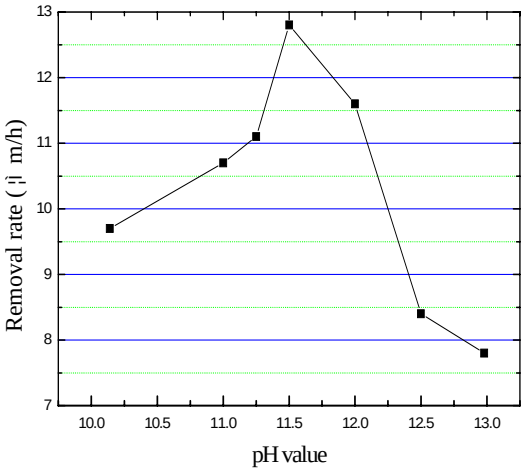
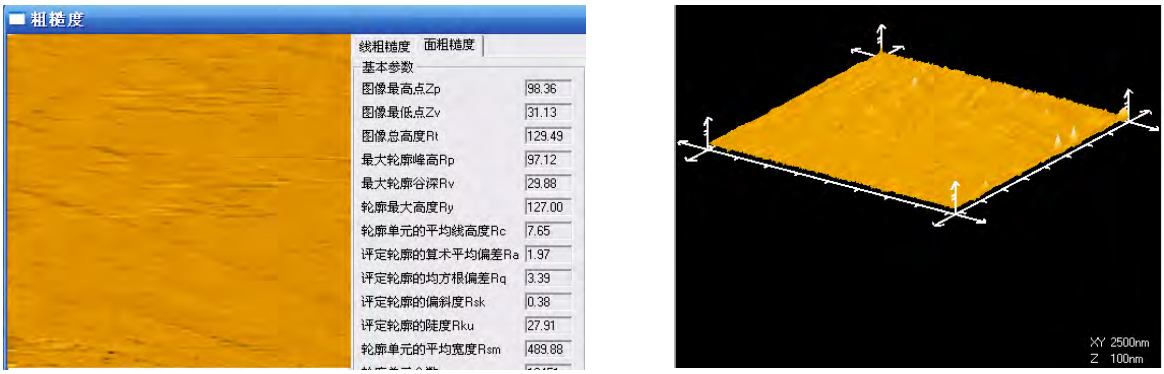


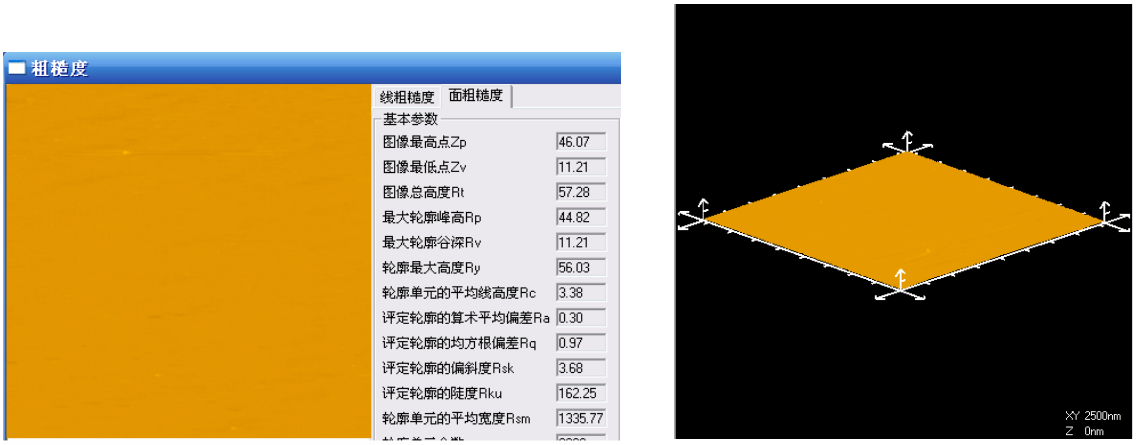
Fig.4 Variation of removal rate as the function of pH value

Optimal Polishing Process

In order to abtain high removal rate and low surface roughness at the whole polishing process on the same polishing machine, the two step polishing mehtod was presented. The aim of first step was to abtain high removal rate. During such step the damage caused by lapping process should be removed and there was no any macroscopic



(a) Pre-polishing（Ra=1.97nm）



(b)Post-polishing（Ra=0.30nm）

Fig.5 AFM surface morphology of the sapphire crystal wafer befor and after CMP（20μm×20μm）

damage seen by microscope. The aim of second step was to abtain low surface roughness. It was to say that under of condition of gaining high removal rate, the pressure and temperature were decreased, slurry flow was increased, and an amount of surfactant was added at ten minutes before the polishing over, which can reduce the surface micro-damage caused by polishing and gain low surface roughness (shown in fig5). Such can provide profitable tech support for industrialization.

Summary

In this study, according to sapphire substrate material property and the polishing mechanism, it was pointed that there was two kinetics process during CMP process,

Ex. 1016, Page 467

which was chemical and mechanical kinetics process respectively. The chemical action is the slow process and takes the control action on the whole kinetics process. According to Arrhenius equation, temperature is the important factor of influencing chemical reaction rate. So the chosen temperature was 45℃. For ensuring the stability of abrasive and chemical action of polishing slurry, the pH value was determined at 11.5~12. In order to abtain high removal rate and low surface roughness at the whole polishing process on the same polishing machine, the two step polishing mehtod was presented. After polishing and cleaning the sapphire surface, the measured removal rate was above 11 μ m /h and the surface roughness was Ra 0.3nm. From the results, it was found that using such method, the optimal sapphire surface can be gotten.

Acknowledgments

The authors would like to thank the Specialized Research Fund for the Doctoral Program of Higher Education of China(20050080007) and Natural Science Foundation of China (No 10676008) for the financial support of this study.

References

1. Wang Chonglu, Sapphire single crystal[M], Tianjin Science and Technology Press, p.16(1983)
2. Jihong Li, Steven R Nutt and Keven W Kirby, J. Am. Cerum. Soc. ,82(11), 3260(1999).
3. Niu Xinhuan, Lu Guoqi and Zhang Weilian, Transactions of Nonferrous Metals Society of China, 16, 187(2006).
4. Wang Yinzhen, Peng Guangliang and Liu Shiliang, Journal of Synthetic Crystals, 34(3), 431,(2005).
5.Honglin Zhu, Luiz A. Tessaroto and Robert Sabia, Applied Surface Science, 236,120 (2004).
6. Tohru Hara, S. Balakumar, Thin Solid Films , 462–463 , 186 (2004).
7. Hong Lei, Jianbin Luo, Wear, 257, 461 (2004).
8. Markus Forsberg, Microelectronic Engineering , 77, 319 (2005).
9. Hahn P O, Microelectronic Engineering, 56, 3 (2001).
10. Wang Xiaohui, Liu Xianglin and Wang Du , Chinese Journal of Semiconductors, 18(11), 867 (1997).
11. Li Bing, Guo Xia and Liu Ying, Semiconductor technology, 39(9), 57 (2005).
12.Yong-Jin Seo and Sang-Yong Kim, Microelectronic Engineering, 65, 371 (2003).
13. Niu Xinhuan, Liu Yuling and Tan Baimei, Transactions of Nonferrous Metals Society of China, 16,732(2006)

ECS Transactions, 18 (1) 441-446 (2009)
10.1149/1.3096483 © The Electrochemical Society

Investigation of Defects in Post-CuCMP

Shiwei Xiong[1], Simon Huang[1], Yifan Yang[1],
Changxing Tan[2], Shan Wang[2], Runtao Zhao[2], Jerry Xu[2]

[1]Wuhan Xinxin Semiconductor Manufacturing Corporation, Wuhan, China
[2]Applied Materials, Shanghai, China

Abstract

In this paper, two particular kinds of defects are investigated. The number of Queue-time defects is strongly impacted by environmental conditions. In the ambient environment, Q-time defects form easily and the number increases greatly with longer queue time. Q-time defect growth can be suppressed indefinitely by storing queued wafers under vacuum or in a nitrogen environment. SEM and EDX analyses show that Q-time defects mostly occur along the Cu line edge and are mainly very small Cu oxide particles. The formation mechanism of Q-time defects is discussed. Another commonly encountered defect – Cu metal missing – which is associated more commonly with previous processing steps, is also discussed.

Introduction

In VLSI circuit fabrication with Cu interconnect technology, post CuCMP defects are always closely monitored [1]. There are many well-known defects associated with CMP tools, such as scratches, Cu or benzotriazole (BTA) residues, slurry-induced corrosion and surface particles. At the same time, there are ongoing efforts to identify new post-Cu CMP defects to understand the root cause and find practical solutions. There are also defects found in post-Cu CMP which are not related to the CMP tool itself but come from non-CMP related sources such as wafer placement, environment conditions and prelayer deposition film properties[2]. These defects should be examined separately under non-CMP related conditions.

In this paper, a particular Q-time defect in post-Cu CMP wafers which is readily influenced by environment conditions will be discussed. We also discuss Cu metal missing defects, another commonly encountered problem in post-Cu CMP wafers. Most of these defects are typically associated with problems that occurred at prior processing steps, such as Cu film quality in the electroplating process.

Experimental

A 90 nm 1P/3M flash memory product was used as the test vehicle. Cu film with thickness of about 0.7 μm was electroplated onto 300 mm wafers after deposition of a Cu barrier (Ta/TaN) and seed layers. Polishing was performed using an Applied Materials Reflexion LK CMP tool using a three-platen configuration: platen one was configured for

Ex. 1016, Page 469

bulk Cu removal, platen two was configured for complete Cu removal at low removal rate and platen three was configured for barrier removal and oxide buffer. The polishing slurries were acid solutions. Post-CMP cleaning was performed using the Reflexion system's Desica Cleaner which included four modules: Megasonic tank, Brush1, Brush2 and Vapor dryer. The key consumables used for post-CMP cleaning were acid solutions for Brush1 and 2, and IPA/N_2 for drying. Defects were inspected with KLA 2367 and Applied Materials' SEMVision G3 STAR tool. The chemical composition of corrosion defects was analyzed by EDX.

To find the relationship between Q-time defect counts and queue time in post-Cu CMP wafers placed in normal environment, samples were placed for different queue times and then inspected every half of an hour from 3 hours to 5 hours. To compare the defect impact of different environments based on the Q-time, several samples were also prepared and placed in vacuum and N2 box with different queue times and then inspected.

Results and Discussions

1. Q-time defect test results and analyses

The relationship between the Q-time defect count and queue time on post Cu CMP wafers placed in the normal fab environment is shown in Figure 1. When queue time was less than 3 hours, Q-time defects were not found. When queue time increased from 3 to 4 hours, Q-time defects gradually increased. SEM review showed that these defects were mainly tiny particles (about 0.1~0.3μm) located on Cu line edges (Figures 2 and 3). EDX analysis showed that chemical composition of the corrosion defects was mainly Cu and O. These defects rapidly increased when queue time was increased to 4.5 hours and longer.

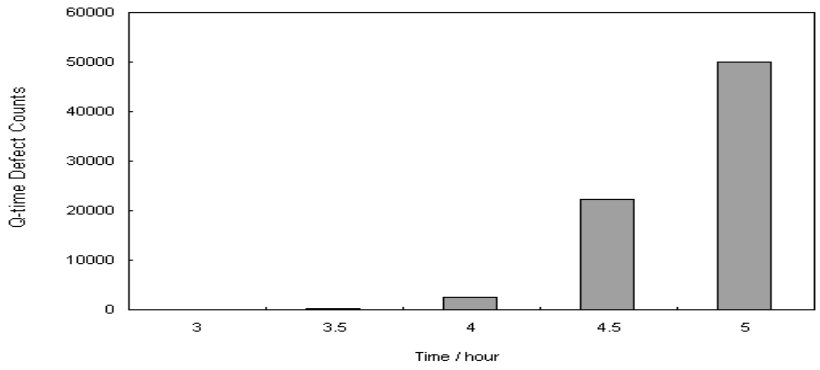


Figure 1. Q-time defect counts versus queue time after CuCMP in normal environment.

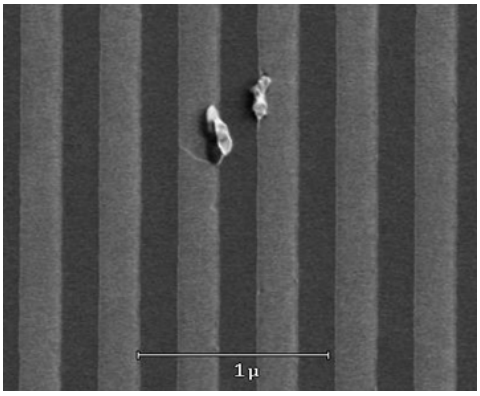


Figure 2. SEM image of Q-time defects

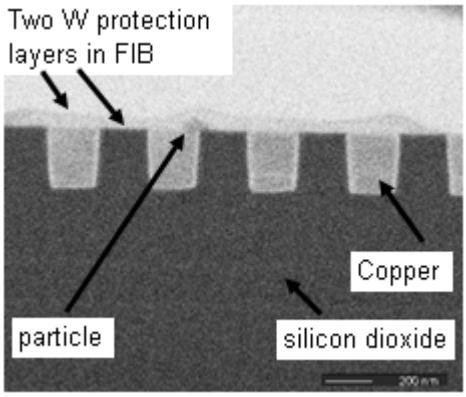


Figure 3. SEM image of inline FIB analysis for Q-time defects.

We also compared the Q-time defect differences when post-Cu CMP wafers were placed under vacuum and in the normal fab environment for 4 hours after complete polishing by a Cu CMP tool. As shown in Figure 4, Q-time defects were not found in vacuum, while ones in a normal environment increased markedly.

The same effect was observed when post-Cu CMP wafers were placed in a nitrogen environment. After 12 and 24 hours Q-time defects could not be found in quantity and overall defect counts did not change (Figure 5). In contrast, Q-time defects increased greatly after placing the wafers in the normal fab environment for 96 hours. Thus, we have demonstrated that Q-time defects of post-Cu CMP wafers are strongly influenced by environmental conditions, while placing the wafers in a vacuum or nitrogen gas environment can avoid Q-time defect formation.

Ex. 1016, Page 471

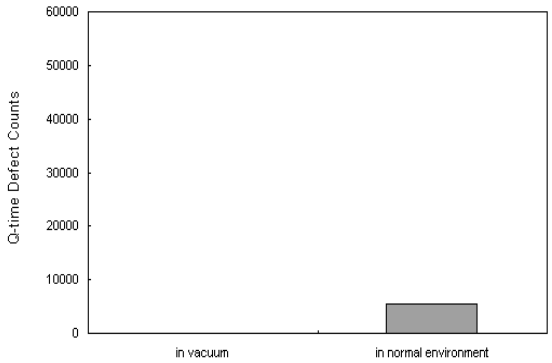


Fig. 4 Comparison of Q-time defect counts for post-Cu CMP wafers placed under vacuum and in normal fab environment for 4 hours.

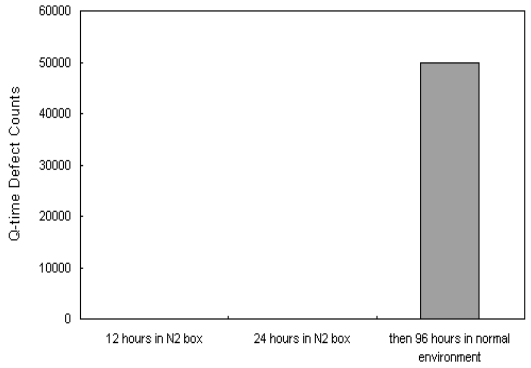


Fig. 5 Comparison of Q-time defect counts for post-Cu CMP wafers placed in nitrogen and normal fab environments.

To understand the root cause of these defects, we focused initially on the CMP tools. However, changing all related consumable parts was not effective in avoiding or decreasing the defect count. Since Q-time defects were readily influenced by the environment, we needed to examine the formation mechanism of Q-time defects based on these conditions.

The normal fab environment includes very active oxygen and Cu that will readily oxidize. After post-CMP cleaning, the wafer surface has a very reactive Cu surface because most compounds such as passivating Cu oxides formed during the polishing process are easily dissolved by the low pH citric acid solution used in post-CMP cleaning [3, 4]. At the Cu line edge, there are also many reactive Cu grain boundaries on the interconnect surface between Cu and its barrier layer. So Cu oxidation can easily occur on the Cu line edge and increase with increasing queue time after post-CMP cleaning [5].

Thus, Q-time defects were mainly caused by environment oxidation. When post-Cu CMP wafers were placed in the normal fab environment, the Cu oxidation would immediately germinate but remain undetectable when the queue time was less than 3 hours. However, with longer queue time, the Cu oxidation would continue until the Q-time defects were easily detectable.

Since vacuum and N_2 gas environments do not contain oxygen, Cu oxidation did not occur. Q-time defects were not found in this environment, even with extended queue time.

Ex. 1016, Page 472

In summary, if the queue time can be kept below 3 hours, the normal fab environment is adequate and the most convenient and cost effective solution. However, when the queue time must exceed three hours, a N2 environment should be used for post-Cu CMP wafers to effectively avoid Q-time defect growth.

2. Cu metal missing defect analyses

In post-Cu CMP wafers, there are also many areas of missing metal that are often associated with previous processing steps, although these defects are usually mistaken as CMP-related pits or corrosion[2]. Here, two kinds of Cu voids were addressed, as shown in Figure 6.

Some of these missing metal areas are formed due to the presence of large rebarbative particles on the wafer surface before the Cu electroplating process. These particles prevent the Cu from completely filling the trench. This kind of large defect is eventually revealed during the Cu CMP process where they appear to be deep, extend across multiple adjacent Cu lines and are irregular in shape (Figure 6 (left)). To prevent this kind of defect, it is critical to remove particles on the wafer surface before completing Cu deposition.

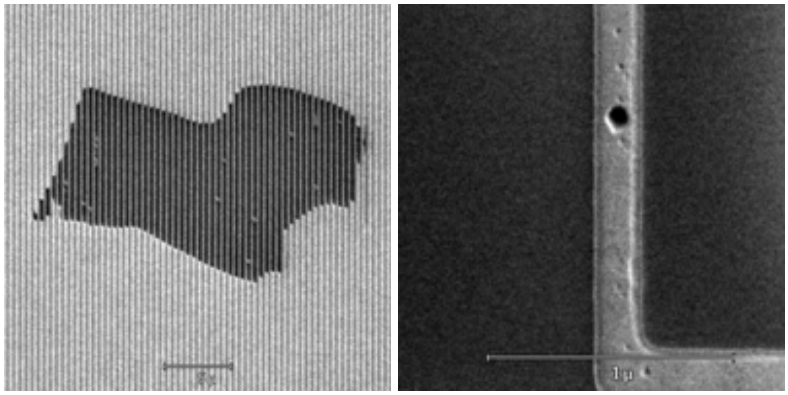


Fig. 6 SEM images of defects caused by missing Cu film

Another kind of missing Cu defect appears as deep voids in the center of Cu lines (figure 6 (right)).This defect is possibly caused by insufficient Cu gap-fill. During film deposition, if key parameters related to the gap-fill process are incorrect, voids can be formed during Cu film deposition. Tuning deposition parameters to increase gap-fill can eliminate this type of defect.

Conclusions

In this paper, it was reported that a particular Q-time defect, unrelated to the CMP tool, was strongly affected by environment conditions. Using a proper queue time or controlling environment conditions was very effective in avoiding this defect type. Defects caused by missing Cu film, associated with previous processing steps, could also be avoided by adjusting prior processing conditions. Thus, when encountering post Cu

Ex. 1016, Page 473

CMP defects, it is important to look beyond the CMP tool at other environmental and pre-processing conditions to identify the root cause of Cu interconnect defects.

Acknowledgments

We would like to thank Mr. Hu Mingjie, Mr. Wang Huaifeng at Wuhan Xinxin Semiconductor Manufacturing Corporation and also Mr. Xu Lai, Mr. Li Hongxin at Applied Material China for their assistance in establishing the CuCMP process.

References

1. Hsueh-Chung Chent, Teng-Chun Tsai, et al., Proc. of IEEE, 6,12 (2001).
2. S. Parikh, J. Educato, A. Wang, et al., Proc. of IEEE, 6, 183 (2001).
3. Y Epshteyn et al., in Proceedings of the CMP-MIC Conference, 323, Tampa (2004).
4. M. Hughes, Naghshineh Shahri, et al., J. Electrochem. Soc., 26, 421(2004).
5. Hsien-Ping Feng, Jeng-Yu Lin, et al., J. Electrochem. Soc., 8, 620(2008).

ECS Transactions, 18 (1) 447-452 (2009)
10.1149/1.3096484 © The Electrochemical Society

Study on the Electro-Chemical Cleaning Technique with BDD Film Anode to Preparation of Electronic Devices

Baohong Gao [a], Yadong Zhu [b], Yuling Liu[a], Xinhuan Niu [a], Jianxin Zhang [c], Xiaoyan Liu [a], Li Li [d]

[a] Department of Information Engineering, Hebei University of Technology, Tianjin,300130,China
[b] School of Electronic Information and Communication Engineering, Tianjin University of technology, Tianjin,300191,China
[c] School of Information and Communication Engineering , Tianjin Polytechnic University, Tianjin,300130,China
[d] School of Electrical Engineering and Automation , Hebei University of Technology, Tianjin,300130,China

Microelectronic devices are the core of the information products, and one of the key points is the cleaning techniques. The cleaning progresses and technology have great influence on the qualities, the price and the competition of the products. The boron-doped diamond (BDD) electrode has a very low capacitance and extreme electrochemical stability, and the BDD electrode was used as anode for ozone generation system through electrolysis of acidic solution. Therefore, the excellent oxidation property of BDD electrode was confirmed and the possibility of further various applications in the cleaning system instead of ozone was expected. This paper presents a new cleaning technology combined electrochemical cleaning, which using BDD film anode with special chemical agents, called electro-chemical cleaning method for short. The principle of the electrochemical progress and some experiment in point was also introduced in this paper.

Introduction

Cleaning techniques is one of the key points in the preparation of micro-electronics devices which is the core of information products. The main contaminations are atoms, ions, molecules and particles adsorbed on the surface of GLSI silicon substrate through physical or chemical adsorption. The requirement of wafer cleaning is to remove impurities without damage on the surface. If one cleaning process is not good, it will affect the rate of excellent products. Affecting the quality, price and competitiveness of products, cleaning techniques and technology plays an important role in science and technology development and social progress. In a sense, no high-quality cleaning, there will be no large-scale integrated circuit or information technology.

At present, the cleaning progress has the highest rate of repeat steps in the integrated circuit chip production, for example, there are 80 multi-cleaning process in the GLSI

Ex. 1016, Page 475

production. Cleaning can be divided into physical and chemical cleaning. How to reduce emissions, to simplify processes or to reduce pollution have not completely solved yet.

In the semiconductor industry，RCA has been widely used. But RCA needs a large number of high-purity chemical reagents (NH_4OH, HCl, H_2O_2, HF, etc.), which increase the cost and the environmental pollution, and it could not meet the need of the development of ULSI. At present, the most popular cleaning method is using HF/O_3, with less chemical reagents, cleaning steps, and achieving better cleansing effect. The water-vapor plasma cleaning and the UV/ozone combined cleaning can also remove effectively organic contaminations on wafers, but both of them have large devices, with complicated steps. As a result, groping for a new wafer cleaning process for the ULSI silicon wafers is imperative.

In this paper, it presents a new cleaning technology combined electrochemical cleaning, which using BDD film anode with special chemical agents, called electro-chemical (the EC-CM) cleaning method for short. The principle of the electrochemical progress and some experiment in point was also introduced in this paper, and the results indicated that the effective cleaning method can remove organic matters, particles and metal ions.

Experiment

The experiment is to show the differences of effects between the RCA and the DE-CM for cleaning silicon substrates

1. RCA Cleaning Method

RCA（SC-1）——$NH_4OH/H_2O_2/H_2O$.Using the weak alkaline NH_4OH to activate the surface of silicon wafer, it will remove particles attached to the surface. In addition, with the volume inverse proportion of NH_4OH,H_2O_2 and H_2O is 1 : 1 : 5, the mixture alkaline cleaning fluid can remove and break down organic contamination at 70 ℃ for 5min.

RCA（SC-2）——$HCl/H_2O_2/H_2O$. With the volume inverse proportion of HCl,H_2O_2 and H_2O is 1 : 1 : 6, the mixture cleaning fluid can remove the metal ions at 70 ℃ for 5min.

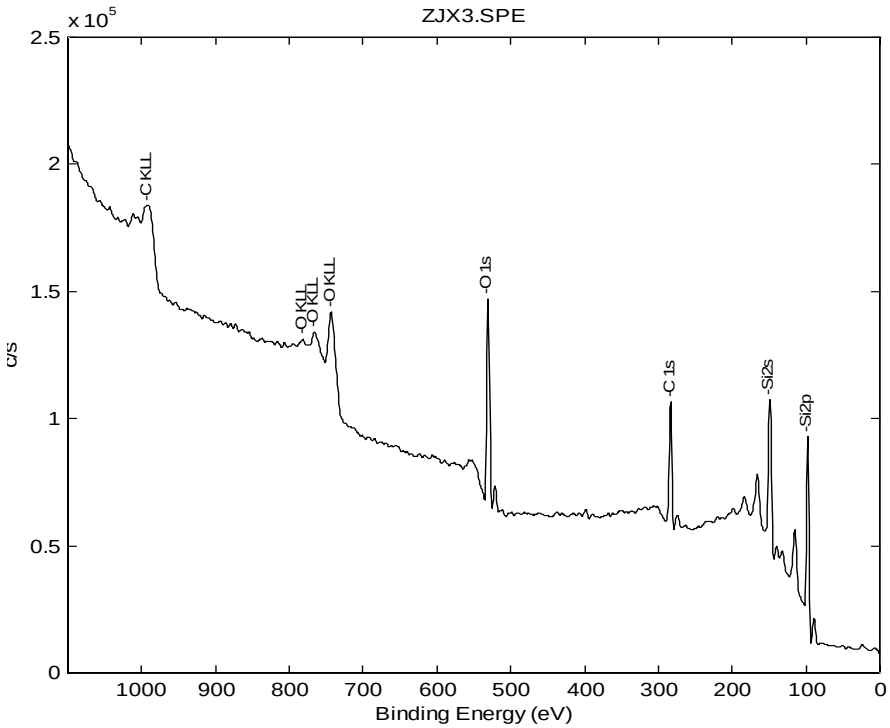
2. The DE-CM

Diamond film electrode electrochemical method with special chemical reagents silicon substrate cleaning, Controling the voltage is 10V, current is 20A, cleaning time is 10min, and the temperature is 70 ℃.

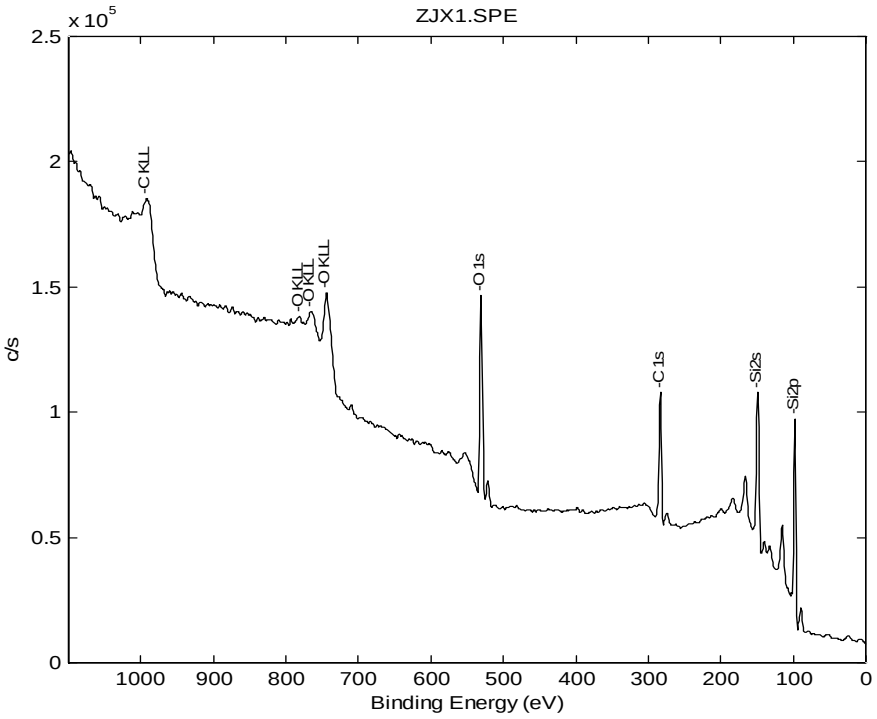
Results and discussion

To confirm the differences of the cleansing effect with the two cleaning method, the two wafers were measured by XPS method, and shown as Figure 1:

Ex. 1016, Page 476



A. RCA Cleaning Progress



B. the DE-CM Cleaning Method

Figure 1 The X-ray photoelectron spectroscopy scanning plans of the surface of the silicon wafers with the two different cleaning progresses.

Figure 1 shows the X-ray photoelectron spectroscopy scans of the two ways to clean the surface of wafer. It clearly shows that the two post-silicon surfaces deled with

different cleaning processes are very similar. To the state, it is constituted mainly by silicon, oxygen and carbon elements.The peak of C1s with binding energy of 284.7eV is the characteristic peak of Organic contamination on the surface, mainly in the form of C-H and C-C bond; The peak of O1s with binding energy 532.52eV is the characteristic peak of SiO_2, indicating the existence of SiO_2 film on the surface of wafer; And the peak of Si2p with the binding energy of 99.50eV is the characteristic peak of single crystal silicon.

Table 1 shows the respective percentage of the three elements calculated from the XPS.

Table1. The concentration of elements on the silicon surface

Elements	RCA cleaning process	the DE-CM cleaning process
C1s concentration percentage	23.16	22.00
O1s concentration percentage	34.18	35.37
Si2pconcentration percentage	42.66	42.63

The above test results show that the wafers cleaned by the two cleaning methods both have a small amount of organic carbon residue and the surfaces have covered a thin SiO_2 film. Comparing the content of elements in the table 1, it shows that the organic carbon residue on the surface cleaned by **the DE-CM** is less than what cleaned by the RCA. The height of SiO_2 peak of the surface cleaned by **the DE-CM** is higher, which shows that **the DE-CM** has a higher capability to remove organic stains on the silicon surface.

Discussion

The EC-CM contains two part: with special active agent，it can remove particles and dissolve organics from the surface of wafer; And using the BDD electrode as the anode, it can produce hydroxyl radical (• OH) and oxygen free radical (• O) (1), which can resolve organics by oxidation. Then the principle will be introduced.

Exclusive surfactant

$$t=Kr^2/f_1f_2S \qquad [1]$$

According to the mathematical model of adsorption priority (equation 1), the exclusive surfactant, a nonionics, can spread wetting with out chemical reaction. It forms a large molecule adsorption priority surface film on the new processed wafer, keeping the surface at the low-energy stability and keeping particles at physical adsorption, which can be easy to remove.

Electrochemical Oxidation

The electrochemical oxidation is processed to dispose the surfactant and organics on the surface of wafers after pulling out of the exclusive surfactant.(2-7)

The equations [2]-[4] show the main electrochemical reaction on the surface of BDD film anode:

Ex. 1016, Page 478

$$H_2O \rightarrow (HO\bullet) + H^+ + e^- \quad [2]$$

• OH, hydroxyl radical, adsorbed on the surface of the BDD film anode that is stable, it will produce oxygen:

$$(HO\bullet) \rightarrow 0.5O_2 + H^+ + e^- \quad [3]$$

Hydroxyl radical (• OH) dissociates into oxygen free radical on the surface of BDD film anode:

$$\cdot OH\text{-}e \rightarrow O\cdot + H^+ \quad [4]$$

The oxidation of the free radicals is very strong, but the lives of them are short. The following equations [5]-[7] show the production of the peroxides, such as hydrogen peroxide and ozone. The oxidation of the peroxide is as strong and more stable.(8)

$$\cdot O + H_2O \rightarrow H_2O_2 \quad [5]$$

$$2\cdot OH \rightarrow H_2O_2 \quad [6]$$

$$3\cdot O \rightarrow O_3 \quad [7]$$

Conclusion

As a new type of cleaning technology, **the DE-CM** combined the specialized reagents and the diamond film electrode electrochemical oxidation, can remove organic residues particles on the surface of wafer. The technology with low cost can achieve excellent cleaning effect and it is simple and easy to use.

References

1. Dae Kyu Kim, Yun Kyu Park, Subhayan Biswas, Chongmu Lee, in Materials Chemistry and Physics. 91, p.490, Removal efficiency of organic contaminants on Si wafer surfaces by the N2O ECR plasma technique. (2005)
2. Samuel Heng, King Lun Yeung, Malik Djafer, Jean-Christophe Schrotter, in Journal of Membrane Science, 289, p.67, A novel membrane reactor for ozone water treatment. (2007).
3. Bogdan C. Donose, Elena Taran, Ivan U. Vakarelski, Hiroyuki Shinto, Ko Higashitani, in Journal of Colloid and Interface Science. 299, p.233, Effects of cleaning procedures of silica wafers on their friction characteristics,(2006).
4. Ahmadi Mohamed Faouzi, Bensalah Nasr, Gadri Abdellatif, in Dyes and Pigments 73, p.86,Electrochemical degradation of anthraquinone dye Alizarin Red S by anodic oxidation on boron-doped diamond, (2007).
5. K. Serrano, P.A. Michaud, C. Comninellis, A. Savall, in Electrochimic Act 48, p431,Electrochemical preparation of peroxodisulfuric acid using boron doped diamond thin film electrodes (2002) .

Ex. 1016, Page 479

6. Madhu Sudan Saha, Tsuneto Furuta, Yoshinori Nishiki, in Electrochemistry Communications 6, p. 201,Conversion of carbon dioxide to peroxycarbonate at boron-doped diamond electrode. (2004).
7. P. Cañizares, M. Arcís, C. Sáez and M.A. Rodrigo, in Electrochemistry Communications, Electrochemical Synthesis of Ferrate using Boron Doped Diamond Anodes. In Press, Accepted Manuscript, Available online 3 July 2007.
8. Fujishima A, Rao T N, Popa E, et al. In Electroanal Chem, p473, Electroanalysis of dopamine and NADH at conductive diamond electrodes (1999).

ECS Transactions, 18 (1) 453-458 (2009)
10.1149/1.3096485 © The Electrochemical Society

Cu/Low-k Thickness Measurement for Advanced Cu CMP Process Development and Control

Y.-L. Li, N. Heylen, T. Delande, K. Kellens,
P. Ong, L.H.A. Leunissen
IMEC, Kapeldreef 75, B-3001 Leuven, Belgium

A. Tarnowka, A. Eliyahu
Nova Measuring Instruments Ltd.

We investigated a wafer level Cu/low-k thickness measurement technique and compared it to the electrical and surface profiling techniques. With this optical technique, we can achieve a comprehensive within-die and within-wafer Cu/low-k thickness monitoring which allows for a more accurate Cu CMP process control in advance Cu/low-k damascene structures.

Introduction

For advanced Cu/low-k interconnect, the normal polishing steps of Cu CMP include bulk Cu polish, Cu clear polish and metal diffusion barrier polish. For both Cu polish steps, there are end point detection algorithms to determine the final polishing times. For barrier polish, the timed algorithm is commonly applied. To guarantee that the whole wafer surface is clear of Cu and barrier residues, over-polish is commonly required for both Cu clear polish and barrier polish. Normally, the low-k dielectric stack and the Cu line in the damascene structures are thick enough to afford a large over-polish process window. However, with the low-k dielectric thickness scaling and as feature sizes shrinking toward the sub-45nm range, the process margin for Cu CMP is shrinking accordingly. More careful tuning of CMP polish recipes is demanding. Since the removal rate depends on quite a few factors, such as the material property, the pattern density, the device layout, etc., and post-CMP cross-section characterization is slow in feedback and the sample size is limited, the demand for a fast, wafer level, in-line measurement of the targeted device structures after polishing is increasing.

A non-destructive low-k thickness measurement technique for patterned wafers will be of significant benefit to the Cu CMP process development and control. Optical measurement, which is non-destructive, fast and sensitive to thin film interfaces, is a promising candidate for such kind of work. NovaScan 3090 is an optical critical dimension (OCD) metrology tool making use of polarized scatterometry. It measures the two polarized signals, transverse electric field and transverse magnetic field, which are then analyzed to retrieve the geometry information of the device structure. In this work, by using NovaScan 3090 integrated in AMAT CMP system Reflexion LK, we investigated the feasibility of wafer level Cu/low-k damascene trench profile characterization.

Experiment

All experiments are based on 300 mm wafers. To explore the material dependency of this optical technology, the test vehicles include two CVD low-k dielectrics integrated in Cu single damascene structures for which the low-k properties are listed in Table I. Figure 1

Ex. 1016, Page 481

shows the schematic cross-sections of the two damascene structures after CMP. For lowk-1, the 13 nm oxide top hardmask could be removed during either etch/strip or CMP. For lowk-2, the 30 nm TiN top hardmask (not shown in the schematic drawing) will be removed during CMP, which requires a longer barrier polish time.

Table I. Properties of the two low-k dielectrics investigated in this paper.

	k value	Porosity	Bottom Hardmask	Initial thickness	Top Hardmask
Lowk-1	3.0	6%	5 nm SiCN/25 nm SiCO	165 nm	13 nm SiO_2
Lowk-2	2.5	25%	35 nm SiCN	180 nm	30 nm TiN

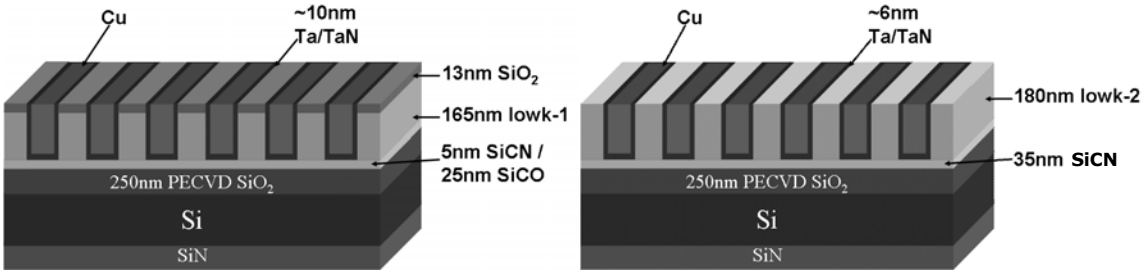


Figure 1. Schematic post-CMP cross-sections of the two Cu/low-k structures.

Cu CMP Polish Conditions

The CMP slurry for Cu clear polish is selective to Cu and the CMP slurry for barrier polish is a non-selective to low-k dielectrics. We modulated the polish times of Cu clear polish and barrier polish steps to get a systematic variation in post-CMP low-k stack thickness from wafer to wafer.

For Cu clear polish, the percentage of over-polish varies from 15% to 75% for both low-k samples. For barrier polish, the polish time for lowk-1 varies from 15s to 25s and the polish time for lowk-2 varies from 37s to 52s. The Cu clear polish and the barrier polish conditions are also shown in Figure 5.

Topography Characterization by HRP

After polishing, the device surface topography was characterized with HRP (High Resolution Profiler). On each wafer, we measured three dies (one at the center, one at the edge and one between the above two) and the target device is a square VDP (Van der Pauw) structure with side length of 80 μm. For the Cu clear polish, a longer over-polish time can introduce more Cu dishing in the VDP structure. For barrier polish, the non-selective slurry has a topography correction capability and thus a longer barrier over-polish time can lead to Cu protrusion. Figure 2 shows the typical HRP traces for Cu dishing and protrusion in the VDP structures.

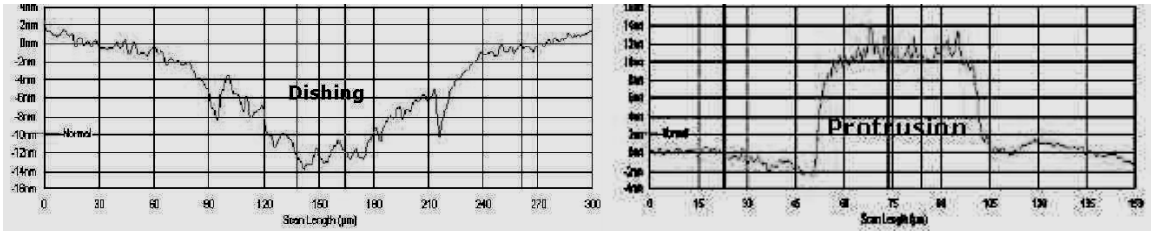


Figure 2. Examples of Cu dishing and sticking out in the HRP traces.

Ex. 1016, Page 482

Wafer Level Optical Characterization

After CMP, we measured all wafers with NovaSCAN 3090 and the target structures are four neighboring meander-forks (as shown in Figure 3). The litho targets of the line-width/spacing are 90/90nm, 95/95nm, 100/100nm and 110/110nm respectively. There are 165 full dies in one wafer and 61 of them are measured.

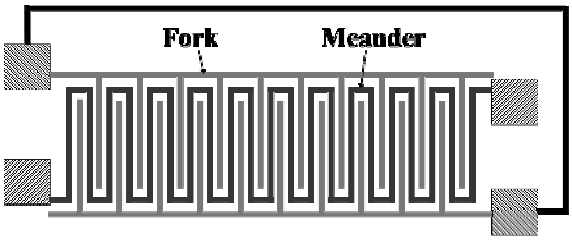


Figure 3. Schematic layout of meander-fork structures (the meander length is 1 cm)

The optical model for geometry fitting is based on a typical single damascene cross-section as shown in Figure 4. The right side of Figure 4 shows the typical fitting curves of the two-dimensional signals.

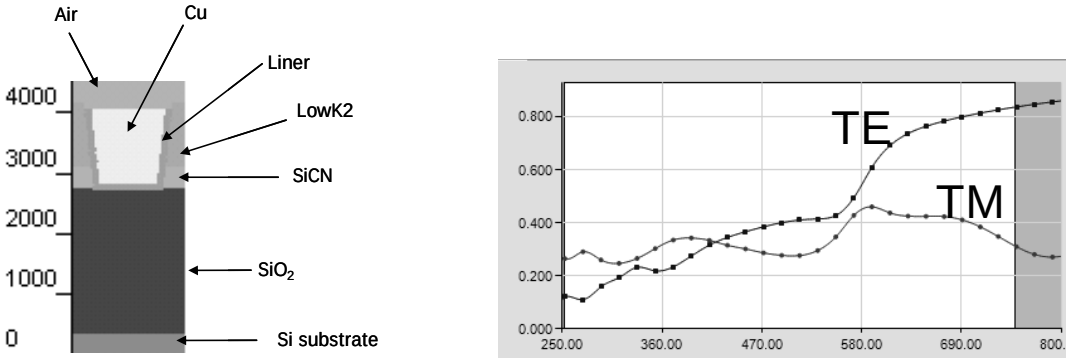


Figure 4. The optical model for geometry fitting (on the left) and the transverse electric field (TE) and transverse magnetic field (TM) fitted with the model (on the right).

Wafer Level Electrical Measurement

Electrical resistance of the meander line is closely correlated to the trench profile. By assuming a quasi-constant Cu resistivity for all line-widths in the target structures, the meander line cross-section area will be linearly proportional to the reciprocal of the metal line resistance, as shown in Equation 1 where R is the metal line resistance, S is the line cross-section area, ρ is the metal resistivity and l is the meander line length.

$$1/R = \frac{S}{\rho \cdot l} \qquad [1]$$

In this experiment, we tested the meander line resistances of the four test structures over the whole wafer.

Result and Discussion

The HRP results for VDP structures are in Figure 5. In the left sides of both figures, the barrier polish time is fixed and the Cu dishing in VDP structures increases with Cu over-polish time. In the right sides of both figures, it is seen that with the Cu over-polish

Ex. 1016, Page 483

setting fixed, a longer barrier polish time with the non-selective slurry can reduce Cu dishing due to the topography correction capability of the non-selective slurry.

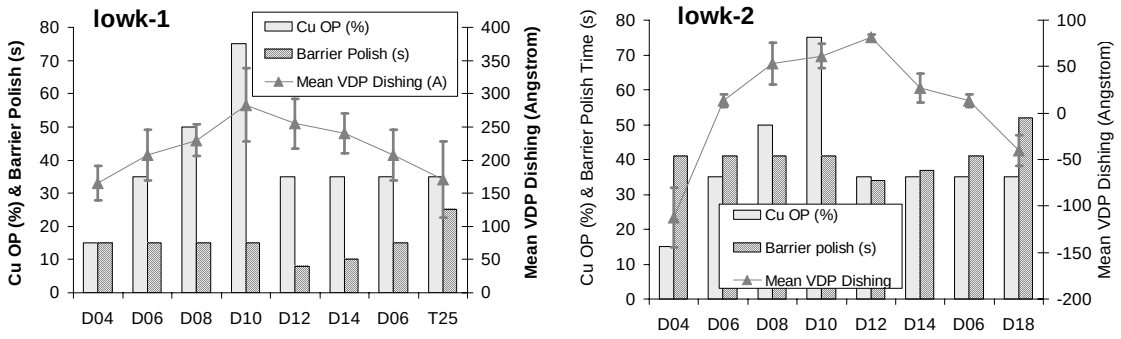


Figure 5. Cu dishing in VDP structures versus Cu over-polish and barrier polish time for Cu/low-k single damascene structures (here negative dishing means Cu protrusion).

Therefore, the designed polish conditions did create different topographies on VDP structures. However, the topography of VDP could be different from those of the meander-forks, because Cu dishing in VDP is much more sensitive to over-polish in Cu and barrier polish steps than low-k erosion in the meander-fork areas.

Figure 6 and Figure 7 show the meander-fork electrical measurement results of both low-k samples. For both low-k dielectrics in Figure 6, when the barrier polish time is fixed, the cross-section area of the meander line decrease with Cu over-polish time, i.e. Cu line thickness shrinks. Figure 7 also demonstrates a similar trend where the Cu over-polish time is fixed and the longer barrier polish time is varied.

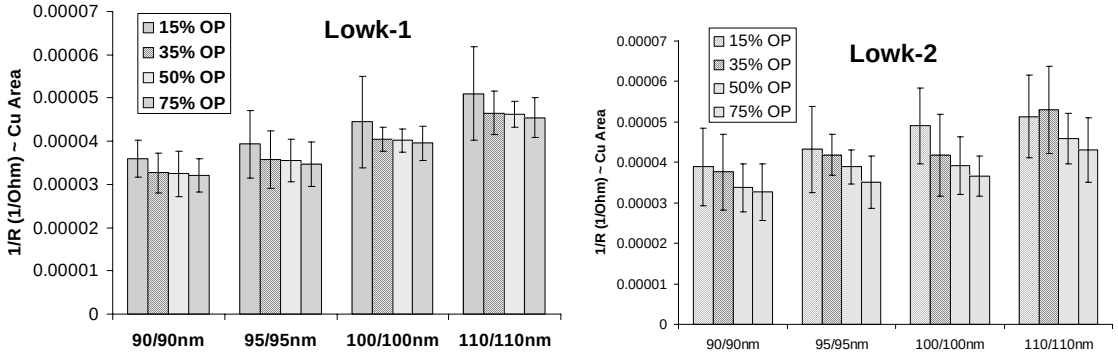


Figure 6. Cu line resistance vs. Cu over-polish at Cu clear step (the barrier polish time is fixed at 15s for lowk-1 and 41s for lowk-2)

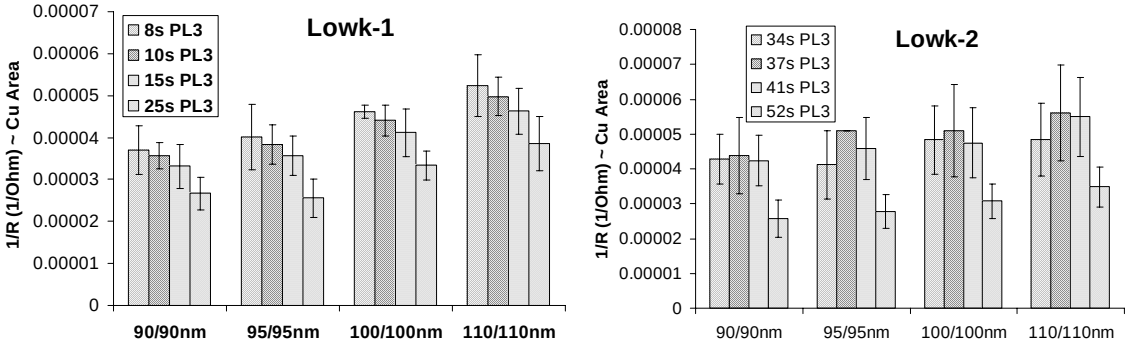


Figure 7. Cu line resistance vs. barrier polish time (Cu over-polish is fixed at 35%)

The averaged Cu line cross-section areas based on optical simulation and the reciprocals of Cu line resistances are plotted together in Figure 8 for the two low-k dielectrics. We can see a systematic match between the optical simulation and the electrical characterization. Since both methods are kind of integration effect, such a good match demonstrates that the optical characterization, based on scatterometry, is an effective technique to characterize post-CMP Cu/low-k damascene structures.

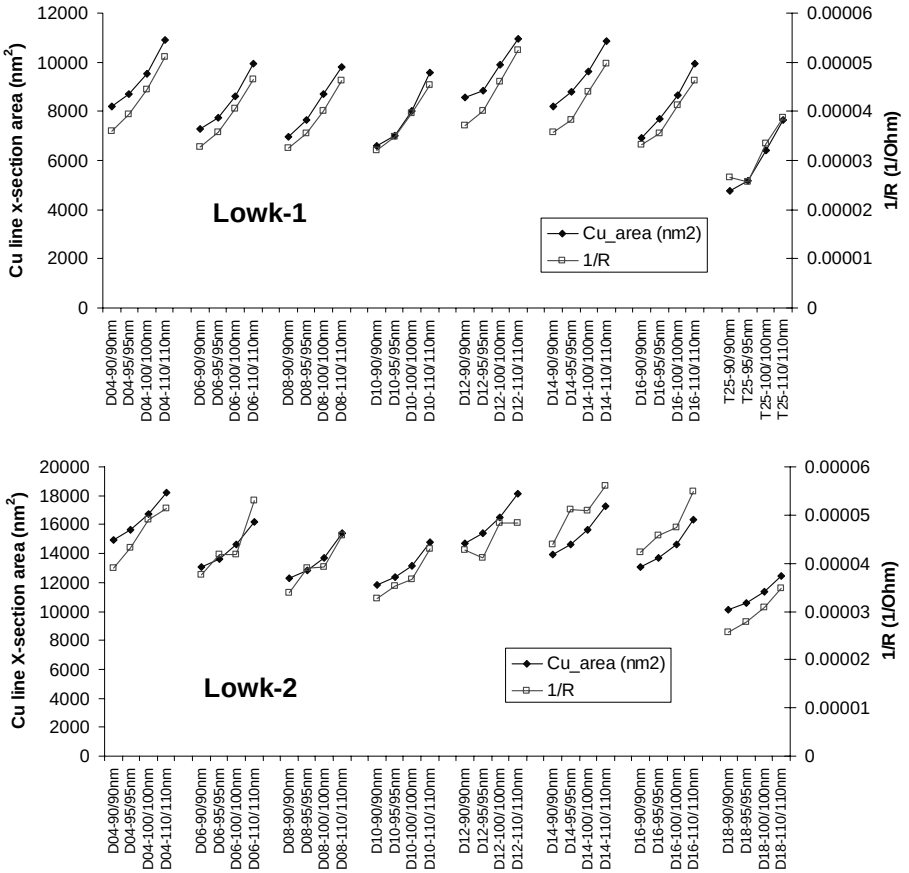


Figure 8. Correlation between the electrical measurement (1/R) and the optically fitted Cu line area.

The advantage of optical characterization is that it is fast, non-destructive and can be integrated into the CMP tool. Moreover, Cu defectivity can influence electrical results but not optical results. For example, NovaSCAN measures a large surface area (a circle around 30um in diameter) and we collect the average CD (critical dimension) and thickness under the measurement spot, whereas Cu defectivity (e.g. Cu oxidation induced voids) can degrade the results of electrical tests. Optical characterization supplies more geometry information than the electrical characterization. Figure 9 shows the low-k thickness and CD across the wafer measured by NovaSCAN.

Combining the slopes of the two curves in Figure 10 and Equation 1, we can calculate the effective resistivity for Cu integrated with lowk-1 and lowk-2. For Cu/lowk-1, the effective resistivity for line widths from 60nm to 100nm is 2.38-μOhm·cm. Considering the Ta based barrier enclosing the Cu line, this is a reasonable result for the effective Cu resistivity on wafer level (1). For Cu/lowk-2, it is 2.98-μOhm·cm, which is apparently

higher than literature. The possible sources for the deviation include Cu oxidation induced voids, line edge roughness etc.

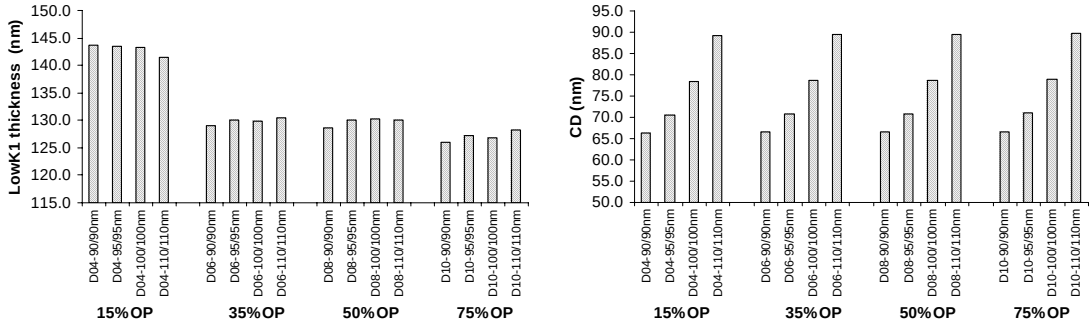


Figure 9. Lowk-1 thickness and Cu line-width (CD) variation with Cu over-polish conditions (15s fixed barrier polish time).

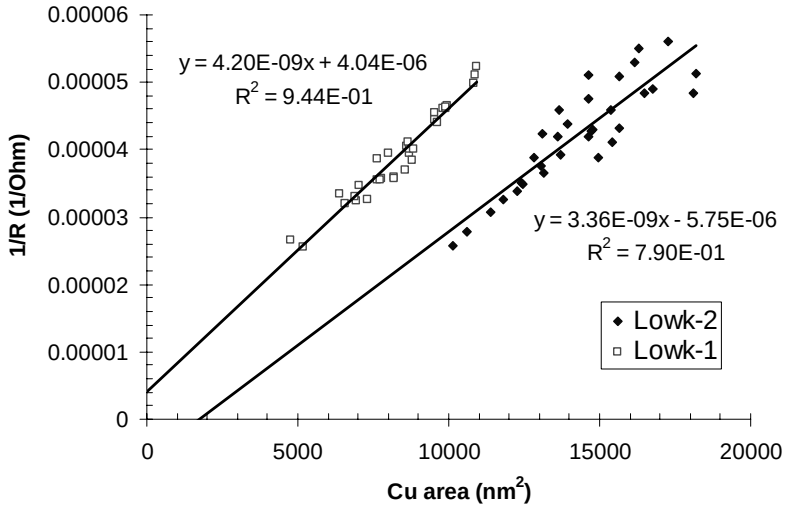


Figure 10. Linear fit between Cu line cross-section areas (from optical simulation) and the reciprocal of the Cu line resistances (from electrical resistance measurements).

Conclusion

The calculated results by optical modeling are in good match with the electrical measurements and fit with the trend of HRP characterization. Longer Cu over-polish induces more Cu dishing and the Cu cross-section area decreases accordingly. Longer barrier polish time can reduce low-k step heights and also reduces Cu cross-section area. By using the optical scatterometry tool NovaSCAN 3090, we have accomplished a comprehensive within-die and within-wafer low-k thickness and Cu profile monitoring which allows for a more accurate Cu CMP process control in advance Cu/low-k damascene structures.

References

1. Plombon *et al*, Appl. Phys. Lett. **89**, 113124 2006

Ex. 1016, Page 486

ECS Transactions, 18 (1) 459-464 (2009)
10.1149/1.3096486 © The Electrochemical Society

Cu/barrier CMP Protection Film Characterization Using Electrochemical Technique

Sunny Xu[a],

[a] Anji Microelectronics (Shanghai) Co., Ltd., Shanghai, 201203, China

As reported in a previous paper, electrochemical data not only can provide the information about Cu corrosion protection, but also can be correlated with slurry CMP performance. In this paper, further study on thermal stability of protection film and Cu planarization efficiency are carried out. Some surface related defect studies correlated with electrochemical data are also discussed.

Introduction

Cu/barrier CMP is commonly used during advanced semiconductor manufacturing process. Electrochemical technique has been widely used in semiconductor manufacturing study, especially in metal CMP processes (1-6). Prof. Srini Raghavan et.al have reported electrochemical study about W & Cu CMP (1-2) and prof. S. V. Babu et.al have reported electrochemical studies about Cu CMP (3-6); yet the correlation between electrochemical data with slurry CMP performance have not been widely reported.

As reported in a previous paper (7), electrochemical data not only can provide the information about Cu corrosion protection, but can also be correlated with slurry CMP performance. Besides those typical performance request for a Cu/barrier slurry, such as material removal rate, WIWNU (within wafer non uniformity), dishing/erosion, ILD loss, Rs distribution control (WIDNU-within die non uniformity), corrosion, scratch, slurry residue, surface defects etc., the slurry thermal stability and Cu planarization efficiency as well as over polish (OP) window are other important aspects which affect greatly the final performance. A successful working Cu/barrier CMP slurry need high planarization efficiency as well as wide over polish window to provide a robust process window. In some cases, when the temperature on pad could increase to 50 °C or even higher, the Cu protection film also need to provide good enough corrosion control to maintain a good surface quality.

In a previous paper, the correlation about material removal rate, protection film formation speed, and dishing with electrochemical data have been discussed. In this paper, further study on thermal stability of protection film and Cu planarization efficiency was carried out. Some surface related defect studies correlated with electrochemical data are also discussed. This is aimed to correlate electrochemical data and CMP process and further more to provide some guideline for slurry and process optimization.

Experimental

Materials and Instrument: Electrochemical set up consists of three electrodes: saturated calomel electrode (SCE) reference electrode, Pt counter electrode, Cu disk electrode (14.5mm Diameter) working electrode and a 100 ml electrochemical cell.

Formulation A, B and C are from Anji microelectronics, Politex pad, table top polisher. Other chemicals are commercially available and used without further purification.

Ex. 1016, Page 487

Electrochemical experiments: working electrode (Cu) was first polished for 1min and then start test in test formulation. Tafel scan potential range was from -1.75 V to 1.75 V and scan speed was 10 mv/s.

Polish process and measurements: 1) Polish with Cu pattern wafer with formulation A, B and C for 1min (Politex Pad), then measure bond pad (80um) dishing with a profiler meter.

Results and discussions

Schematic Illustration for Cu/barrier CMP with Electrochemical Data

Cu planarization mechanism can be understood as preferential removal of protruding area over the recessed area for given pattern feature. Ideally, there will be no dishing after reaching planarization. In reality dishing/protrusion is always present. The root cause for dishing generally includes two parts: mechanical part and chemical parts. For given polishing parameters (down force, pad, rotation speed, flow rate etc..), the mechanical part will be more or less fixed; yet the chemical wet etch effect on the protruding area and recessed area is not the same, different slurry system resulting different dishing performance are mainly due to different protection mechanism. This study will mainly focus on the chemical aspect on dishing performance assuming the mechanical parts are similar. Electrochemical data with and without polishing can be used to illustrate the protection film related performance (such as dishing, planarity efficiency, OP window, corrosion related surface defects etc.). This can be schematically illustrated in Fig. 1.

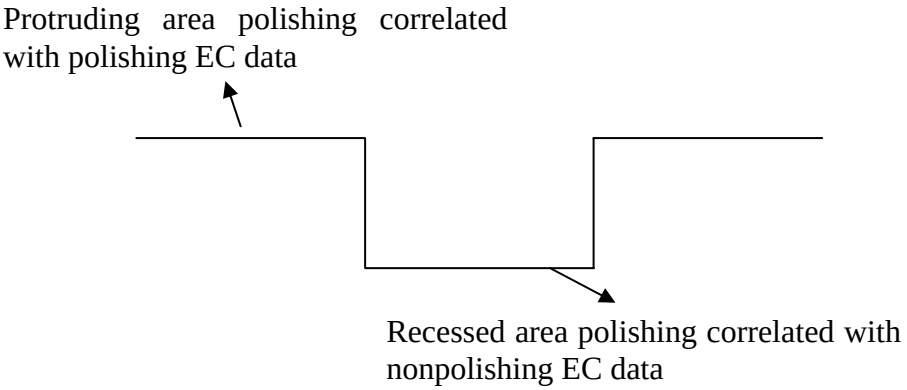


Fig. 1 Schematic illustration for Cu/barrier CMP with electrochemical data

As Fig. 1 shown above, it is easily understood that the local down force applied on the recessed area will be lower than that of the protruding area. This difference helps us to understand the Cu planarization and dishing control mechanism. Fast protruding area removal provide sufficient material removal rate while good recessed area protection contributes a good dishing control. The recessed area protection was mainly through good protection film protection during the process. Based on this understanding, tafel plot technique was used to correlate with the actual CMP process. Electrochemical data with polishing can provide the information about protruding area polishing while the tafel plot without polishing provides the information about the recessed area polishing. The combination of these two sets of data can provide the information for Cu removal rate and dishing control and also provide information about planarity efficiency.

Ex. 1016, Page 488

Electrochemical Technique for Cu Protection Film Thermal Stability Study

In Cu/barrier CMP process, the temperature on pad could easily raise, in some cases, it can reach 50 degree or even higher (see Fig. 2A). After long time continuous polishing, the temperature on polishing pad is not the same as before, this is one of the reasons why the corrosion related problem is becoming worse and more frequently happened after long time CMP processes. The protection film thermal stability could help to indicate if the protection film is stable enough to protect Cu film in such exhausting condition. Generally the higher the dishing the weaker the protection film is. The Cu bond pad dishing is then used as an indicator for Cu protection film thermo stability in the following discussion. Tafel plot is one of the most commonly used electrochemical techniques for metal corrosion study. In our experiment, corrosion potential and corrosion current were used to correlate with the Cu bond pad dishing in different temperature. Corrosion current increase indicating protection film is less stable in such temperature. In general the higher the corrosion current is, the less stable the protection film will be. For the same formulation in different temperature, our correlations were listed below:

TABE I. **Cu Protection Film Thermo Stability Study**

Temperature(°C)	18	28	38	48
Ecorr (mV vs.)	351	361	369	378
Icorr (uA)	4.48	4.22	6.54	7.86
Dishing (A)	560	430	650	860

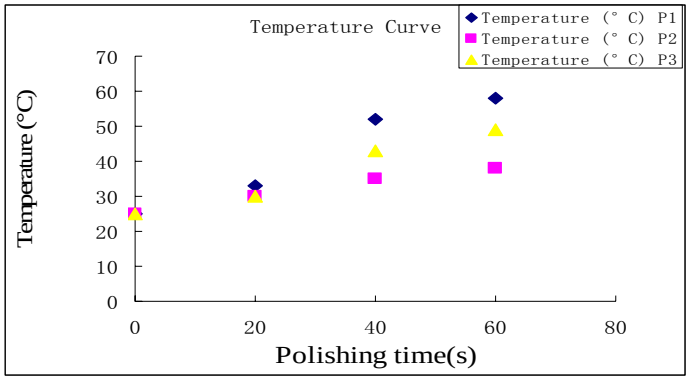


Fig. 2A Temperature curve for Cu CMP during polishing

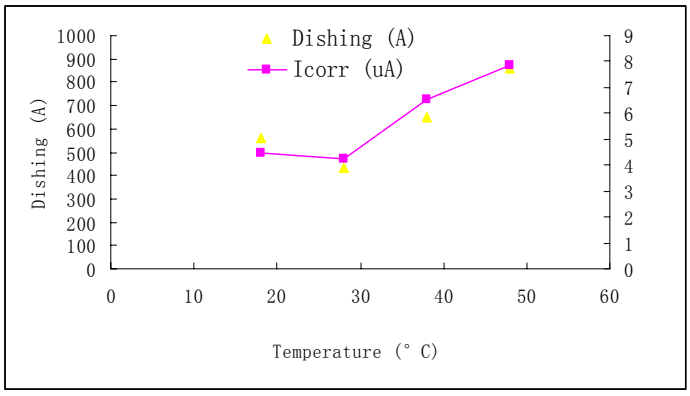


Fig. 2B Correlation between electrochemical data and Cu bond pad dishing

Electrochemical data and Cu bond pad dishing show good correlation in Fig.2B. In the same formulation system, the Cu protection film stability depends on the competition reaction of its chemical composition on Cu surface, including the complexing agent, oxidizers and inhibitors. CMP process is trying to find the favorable balance among them. When temperature change, the original balance is broken, the temperature increase could enhance the complexing agent and oxidizers chemical reaction rate with Cu surface, Also, it could increase the inhibitor reaction with Cu, which one plays more important role in the CMP process will finally result in the protection film stable or not. Electrochemical study for the protection film thermo stability shows a "V" shape trend in the existing slurry system which indicating there is competing process during temperature increase, the inhibitors effect play more important role when temperature around 28°C while the complexing agent and oxidizer play more important role in other temperatures, this could help to optimize the process condition.

Electrochemical Technique for Cu Planarity Efficiency and OP Window Study

A successful working Cu/barrier CMP slurry need high planarity efficiency as well as wide over polish window to provide a robust process control. For different protection film, Cu planarity efficiency and OP window have significant different. There is a balance between the planarity efficiency and OP window. Ideally, fast planarity efficiency and wide OP window is always preferable.

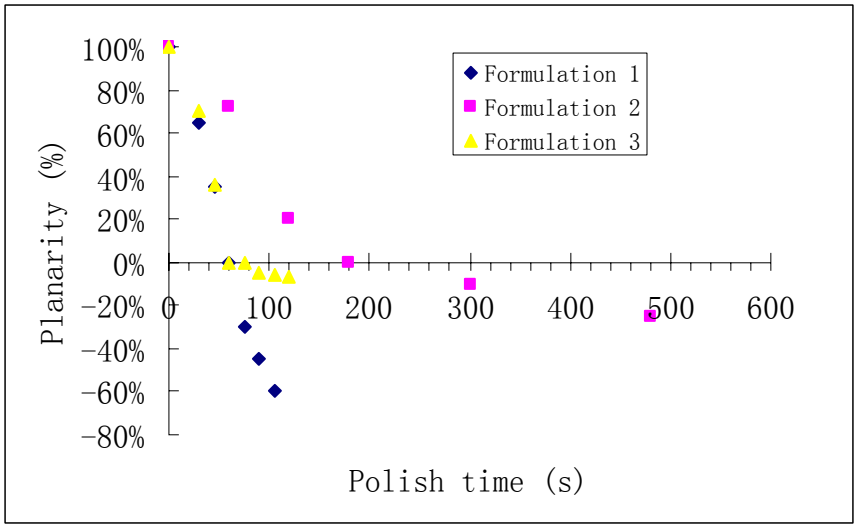


Fig. 3 Cu planarity efficiency and OP window study in different formulation system

In Fig. 3, three different formulation systems were investigated. Formulation 1 has the best planarity efficiency but poor OP window; Formulation 3 has poor planarity efficiency and good OP window; Formulation 2 is kind of optimized formulation, which shows good planarity efficiency and reasonable OP window. For the same three formulations, the electrochemical data were also plotted in Fig. 4, the anodic current and anodic potential were plotted, and it shows the similar trend as Fig.3.

Ex. 1016, Page 490

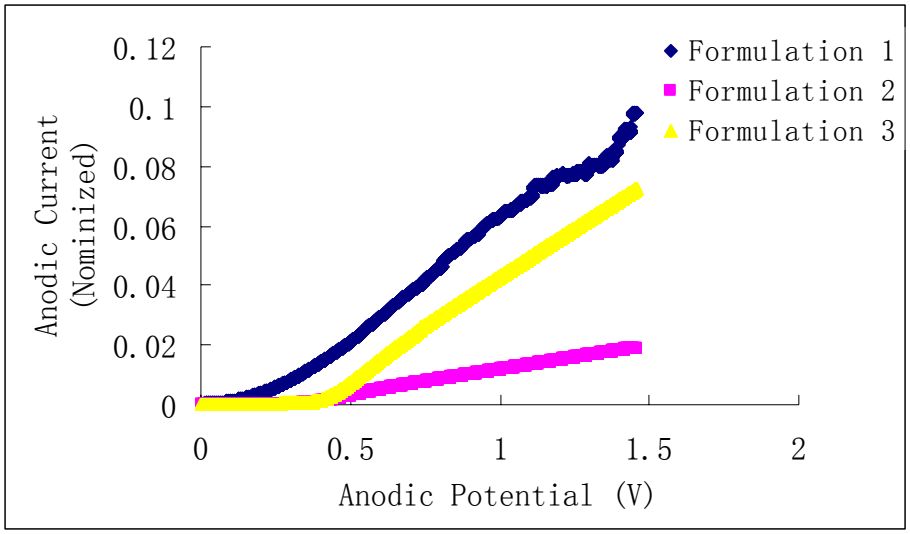


Fig. 4 Cu protection film electrochemical study in different formulation system

For formulation 1, the current increases very fast when the voltage goes higher, indicating there is not good protection film on the Cu film; Formulation 3 shows the current increase quite slow when the potential increases, indicating a strong protection film, which may lead into small removal rate; Formulation 2 reaches a good balance between removal and protection and finally shows a good planarization efficiency and reasonable wide OP window.

Electrochemical Technique for Corrosion Related Cu Surface Defect Study

Cu surface defect is another big challenge for Cu/barrier CMP process. In this study, we mainly focus on corrosion related surface defects; since it is found that the nature of protection film has significant impact on surface defects.

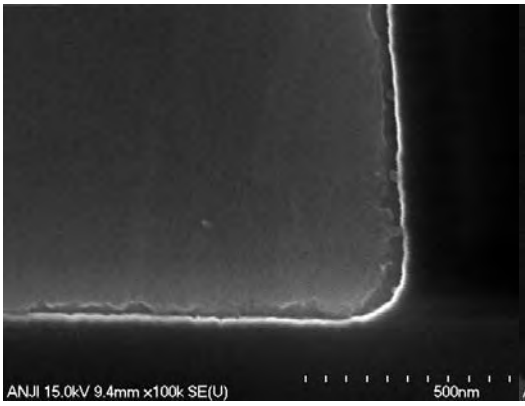


Fig.5A Formulation 1

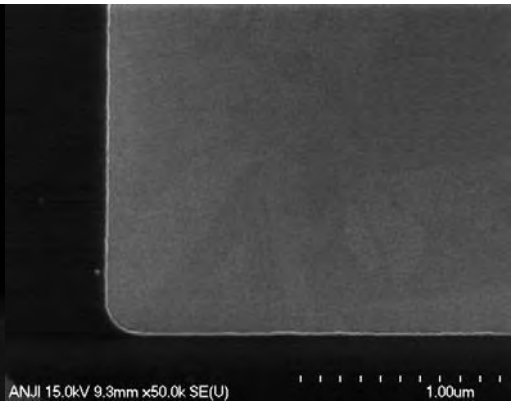


Fig.5B Formulation 2

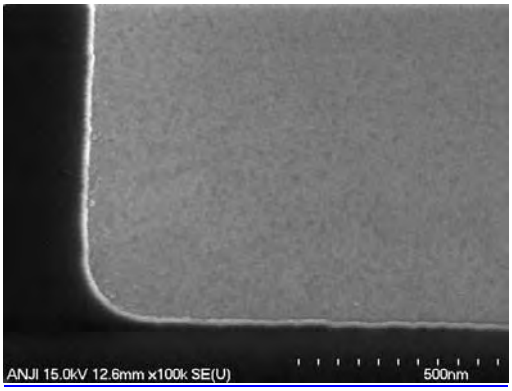


Fig.5C Formulation 3

Fig. 5 SEM picture for different slurry system after CMP

As Fig 5 shows, different protection film shows different surface finish after the same CMP process. As discussed in Fig. 4, formulation 1 has a poor protection film and shows clearly edge corrosion after polish; formulation 3 show strong protection and looks like there is some residue like structure remaining on the Cu surface, formulation 2 is the optimized formulation which shows the best surface finish.

Conclusions

Electrochemical data can not only provide the information for Cu corrosion protection, but it can also be correlated with slurry CMP performance. The protection film thermal stability and planarization efficiency and OP window study were correlated with some electrochemical data. Some surface related defects were also discussed, It shows electrochemical technique is a good method to reach some fundamental understanding of slurry system and can provide good help for slurry and process optimization.

Acknowledgments

The author would like to thank members from R&D Department at Anji Microelectronics (Shanghai) Co., Ltd. for their kind support.

References

1. Emil A. Kneer, Chilkunda Raghunath, and Srini Raghavan, *J. Electrochem. Soc.*, **143**, No. 12, (1996)
2. Subramanian Tamilmani, Wayne Huang, Srini Raghavan, and Robert Small, *J. Electrochem. Soc.***149** (12) G638-G642, 2002
3. Udaya B. Patri, Serdar Aksu, and S. V. Babu, *J. Electrochem. Soc.* **153**, G650 (2006)
4. Y. Hong, V. K. Devarapalli, D. Roy, and S. V. Babu, *J. Electrochem. Soc*. **154**, H444 (2007)
5. V. R. K. Gorantla, K. A. Assiongbon, S. V. Babu, and D. Roy, *J. Electrochem. Soc*. **152**, G404 (2005)
6. Jindal and S. V. Babu, *J. Electrochem. Soc*. **151**, G709 (2004)
7. Sunny Xu and Shumin Wang, *ISTC2008 proceeding* 526 (2008)

Ex. 1016, Page 492

ECS Transactions, 18 (1) 465-471 (2009)
10.1149/1.3096487 © The Electrochemical Society

Effect of Polisher Kinematics in Reducing Average and Variance of Shear Force and Increasing Removal Rate in Copper CMP

Yasa Sampurno [a,b], Ara Philipossian [a,b], Sian Theng [a,b], Takenao Nemoto [c], Xun Gu [c], Yun Zhuang [a,b], Akinobu Teramoto [c] and Tadahiro Ohmi [c]

[a] Araca, Inc., 2550 East River Road, Suite 12204, Tucson, Arizona 85718, USA
[b] University of Arizona, 1133 James E. Rogers Way, Tucson, Arizona 85721, USA
[c] Tohoku University, 6-6-10, Aza-Aoba, Aramaki, Aoba-ku, Sendai 90-8579, Japan

The effect of polisher kinematics on average and variance of shear force and removal rate in copper CMP is investigated. A 'delamination triangle' consisting of average shear force, variance of shear force, and required polishing time is defined, and 'delta' is calculated based on the product of the above three components. In general, low values of 'delta' are preferred to minimize defects during polishing. In the first part of this study, 200-mm blanket copper wafers are polished at constant platen rotation of 25 RPM and polishing pressure of 1.5 PSI with different wafer rotational rates and slurry flow rates. Results indicate that at the slurry flow rate of 400 ml/min, 'delta' is higher by 50 to 290 percent than at 200 ml/min, and increasing wafer rotational rate from 23 to 148 RPM reduces 'delta' by more than 90 percent and improves removal rate within-wafer-non-uniformity by 2X. In the second part of this study, polishing is performed at the optimal slurry flow rate of 200 ml/min and wafer rotational rate of 148 RPM with different polishing pressures and platen rotational rates. Results indicate that 'delta' is reduced significantly at higher ratio of wafer to platen rotational rates.

Introduction

During CMP, 2-body and 3-body interactions among the wafer, the slurry particles and the pad generate a wide range of shear forces. With IC devices shrinking further into nano-scale dimensions and low-k nano-porous materials becoming increasingly attractive, shearing needs to be re-visited vis-à-vis the mechanical integrity of the metal and underlying porous dielectric stack since high average shear force, in combination with its large fluctuating component and long polishing times (especially to clear the barrier) become major defect-causing problems. In this paper, the contributions of shear force (φ), variance of shear force and required polishing time are presented independently in a 3-axis graph called the 'delamination triangle'. The value of 'delta' is calculated based the product of the above three components and normalized to illustrate the effect of polisher kinematics during copper CMP process.

Ex. 1016, Page 493

Experimental Apparatus and Procedure

All polishing experiments were performed on an Araca APD-800 polisher and tribometer. The polisher is equipped with the unique ability to acquire shear force and down force in real-time.[1] In this study, force acquisition rate was set at 1,000 Hz. An IC1000 pad with Suba IV sub-pad was used. The pad had positive 20° slanted concentric and 0° logarithmic positive groove design. A 100-grit triple ring dots (TRD) diamond disc from Mitsubishi Materials Corporation was used to condition the pad at 5.8 lb_f during wafer polishing. The conditioner disc rotated at 95 RPM and swept 10 times per minute. During polishing, the diamond disc, pad and wafer rotated in the counter-clockwise direction. Hitachi Chemical HS-2H-635-12 slurry was used. The polishing time for each wafer was 90 seconds. In the first part of the study, the effect of slurry flow rate and wafer rotational rate was explored at constant platen rotation of 25 RPM and polishing pressure of 1.5 PSI. The slurry flow rate was varied at 200, 300 and 400 ml/min. The wafer rotational rate was varied at 23, 98 and 148 RPM. In the second part of the study, the effect of polishing pressure and platen rotation rate was explored using the optimal slurry flow rate (200 ml/min) and wafer rotational rate (148 RPM) obtained from the first part of the study. The polishing pressure was varied at 1.5, 2 and 2.5 PSI. The platen rotational rate was varied at 25, 40 and 55 RPM.

Delamination Triangle

During CMP, 2-body and 3-body interactions among the wafer, the slurry particles and the pad result in polished substrate removal and generate a wide range of shear forces. In general, higher average shear force and variance of shear force, as well as longer polishing time contribute to delamination during polishing. In this paper, the contributions of average shear force (φ), variance of shear force (σ^2) and required polishing time (τ) are presented independently in a 3-axis graph called the 'delamination triangle'. The value of 'delta' (Δ) which is the product of the above three components represents the combined contributions in the 'delamination triangle'. Figure 1 shows a 'delamination triangle' corresponding to a 5-minute polishing with an average shear force of 12 lb_f and shear force variance of 8 lb_f^2. The 'delta' value is 480 $min_x lb_f^3$.

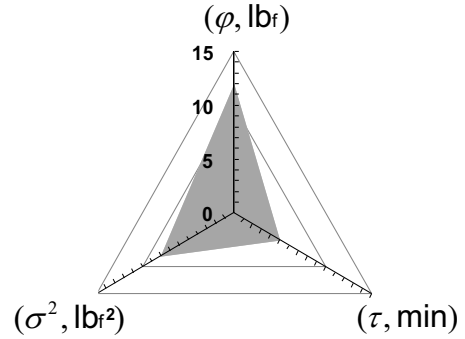


Figure 1. Example of a 'delamination triangle'

The 'delamination triangle' can be normalized by dividing the original value by a reference value for each axis. A reference process can be selected and Figure 2(a) shows the normalized 'delamination triangle' for a reference process. For the reference process,

the value of each axis is unity in the normalized 'delamination triangle'. Therefore, the value of normalized 'delta' is also unity. If the example shown in Figure 1 is used as the reference process, Figure 2(b) shows the normalized 'delamination triangle' corresponding to a 2-minute polishing process with an average shear force of 7.2 lb_f and variance of shear force of 4 lb_f^2. The normalized value in the 3-axis are 0.6, 0.5 and 0.4 for the average shear force, variance of shear force, and required polishing time, respectively. The normalized 'delta' value is 0.12.

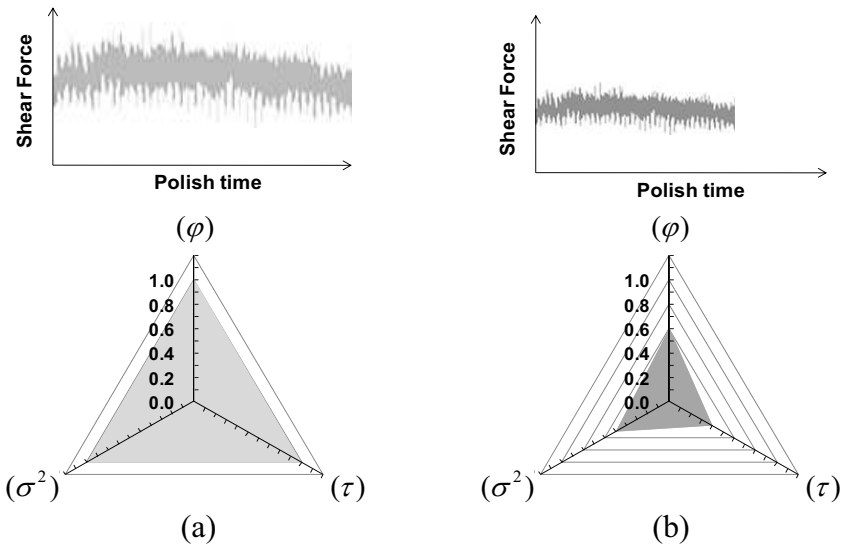


Figure 2. Normalized 'delamination triangle' corresponding to (a) baseline process and (b) improved CMP process

Asynchronous Pad and Wafer Rotation Rate

In a wafer-centered polar coordinate, the pad-wafer relative velocity, $V_{P\text{-}W}$, at any point on the wafer (r,Θ) differs from the sliding velocity at the wafer center $(r=0)$ by a factor $\ddot{A}$ that depends on the wafer to platen rotation rate ratio Ω_w/Ω_p and the pad-wafer center separation R as follows: [2]

$$\ddot{A}(r,\theta) = \frac{V_{P-W}(r,\theta)}{V_{P-W}(0,\theta)} = \sqrt{1 + 2\left(1 - \frac{\Omega_W}{\Omega_P}\right)\left(\frac{r}{R}\right)cos\,\theta + \left(1 - \frac{\Omega_W}{\Omega_P}\right)^2\left(\frac{r}{R}\right)^2} \quad [1]$$

For synchronous pad-wafer rotation, the factor $\ddot{A}$ is unity at any point on the wafer (r,Θ). For asynchronous pad-wafer rotation, the average asynchronous pad-wafer sliding velocity, A, across the wafer radius R_w can be calculated based on the area weighted distribution as follows:

$$A = \frac{\int_0^{2\pi}\int_0^{R_W} \ddot{A}(r,\theta) r\, dr\, d\theta}{\int_0^{2\pi}\int_0^{R_W} r\, dr\, d\theta} \quad [2]$$

Table 1 summarized the average asynchronous pad-wafer sliding velocity, A, for five combinations of Ω_w and Ω_p used in this study. The value of A was calculated for 200-mm wafers with pad-wafer center separation, R, of 225 mm.

Ex. 1016, Page 495

Ω_p	Ω_w	A
25	23	1.00
25	98	1.21
25	148	1.62
40	148	1.18
55	148	1.07

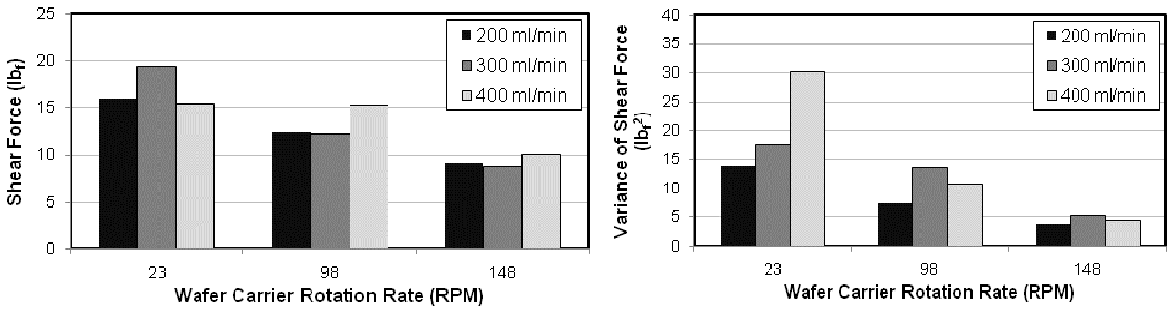
Table 1. The average asynchronous pad-wafer sliding velocity, A

Results and Discussion

Effect of slurry flow rate and carrier rotation rate

In the first part of the study, the effect of slurry flow rate and wafer rotational rate is studied at a constant platen rotation of 25 RPM and polishing pressure of 1.5 PSI. This part of the study aims to select the optimum slurry flow rate and wafer rotation rate to achieve lower 'delta' values and better removal rate within-wafer-non-uniformity (RR WIWNU).

Figure 3(a) and 3(b) shows the average and variance of shear force as a function of wafer rotational rate and slurry flow rate. Results indicate that increasing the wafer rotation rate from 23 RPM to 148 RPM reduces the average and variance of shear force by approximately 40 percent and 80 percent, respectively. In comparison, there is no consistent trend in the effect of slurry flow rate on the average and variance of shear force.



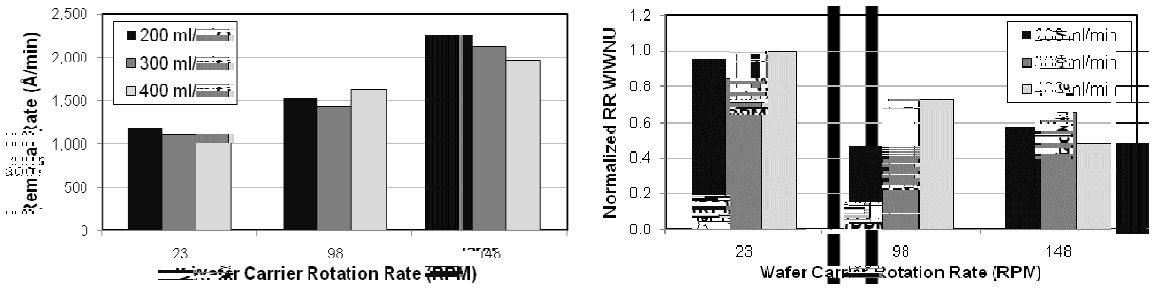
(a) (b)

Figure 3. (a) Average shear force and (b) variance of shear force as a function of wafer carrier rotation and slurry flow rate

Figure 4(a) shows that higher wafer rotational rates lead to higher copper removal rates. As indicated in Table 1, at a constant platen rotational rate, higher wafer rotational rates result in higher sliding velocities, leading to higher removal rates. Figure 4(b) shows the normalized RR WIWNU as a function of wafer rotational rate and slurry flow rate. The RR WIWNU at wafer rotational rate of 23 RPM and slurry flow rate of 400 ml/min is employed as a baseline for normalization. Increasing the wafer rotational rate from 23 RPM to 148 RPM reduces the RR WIWNU by approximately 45 percent. At wafer and platen rotational rate of 23 RPM and 25 RPM, removal rate profile indicates a center-fast removal mechanism. Increasing the wafer rotational rate leads to higher local sliding

Ex. 1016, Page 496

velocity toward the wafer edge, enhances the local removal rate, and improves the RR WIWNU. In comparison, there is no consistent trend in the effect of slurry flow rate on the removal rate and RR WIWNU.

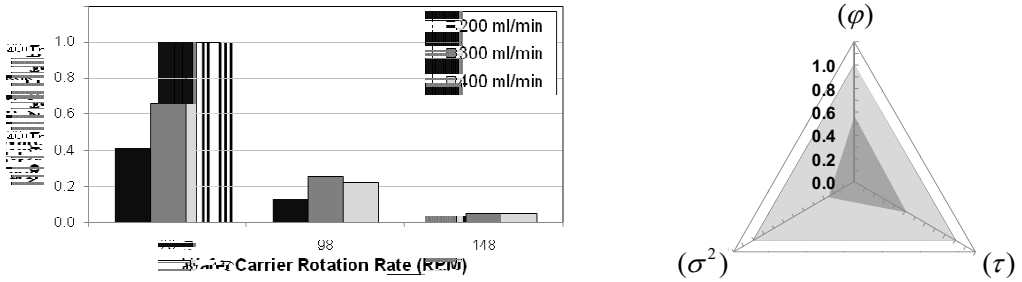


(a) (b)

Figure 4. (a) Removal rate and (b) normalized RR WIWNU as a function of wafer carrier rotation and slurry flow rate

Figure 5(a) shows the normalized 'delta' values. The process at the wafer rotational rate of 23 RPM and slurry flow rate of 400 ml/min is used as the reference process for normalization. The results indicate that the 'delta' values are 91 to 96 percent lower at the wafer rotational rate of 148 RPM compared with at 25 RPM. Figure 5(a) also indicates that the 'delta' values increase with the slurry flow rate. Figure 5(b) shows the normalized 'delamination triangle' for the process with the lowest 'delta' value. The lowest 'delta' value is obtained when the slurry flow rate was at 200 ml/min and the wafer rotational rate was at 148 RPM. Under this condition, the removal rate within-wafer-non-uniformity improves by a factor of 2 compared to the reference process.

Therefore, increasing the ratio of Ω_w/Ω_p leads to significantly lower values of 'delta'.



(a) (b)

Figure 5. (a) Normalized 'delta' as a function of wafer rotational rate and slurry flow rate; (b) normalized 'delamination triangle' at polishing pressure of 1.5 PSI, slurry flow rate of 200 ml/min and Ω_p/Ω_w of 25/148 RPM

Effect of polishing pressure and platen rotation rate

In the second part of the study, the effect of polishing pressure and platen rotational rate is explored under the optimal slurry flow rate (i.e. 200 ml/min) and wafer rotational rate (i.e. 148 RPM) obtained in the first part of the study. Table 2 summarizes the five

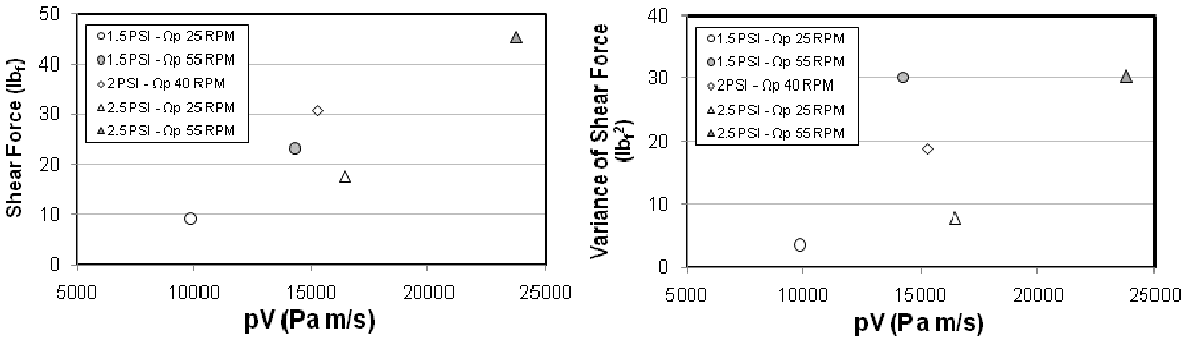
Ex. 1016, Page 497

different combinations of polishing parameters used in this part of study. The No. 1 polishing parameter shown in Table 2 achieves the lowest 'delta' value in the first part of the study.

No.	Pressure (PSI)	Ω_p (RPM)	Ω_w (RPM)	Slurry Flow Rate (ml/min)
1	1.5	25	148	200
2	1.5	55	148	200
3	2	40	148	200
4	2.5	25	148	200
5	2.5	55	148	200

Table 2. Polishing conditions to explore the effect of polishing pressure and platen rotation rate

Figures 6(a) and (b) show a non-linear trend in the average and variance of shear force as a function of the product of polishing pressure and sliding velocity (pV). At 2.5 PSI and platen rotational rate of 25 RPM, the average and variance of shear force decrease significantly from the trend.

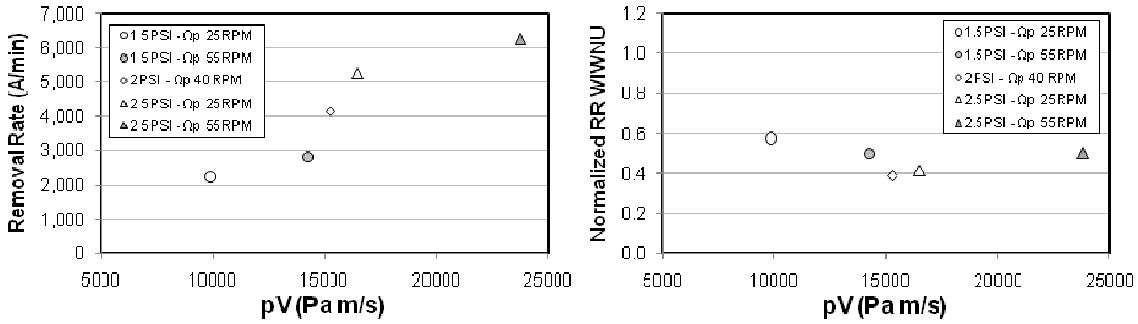


(a) (b)

Figure 6. (a) Average shear force and (b) variance of shear force as a function of pV

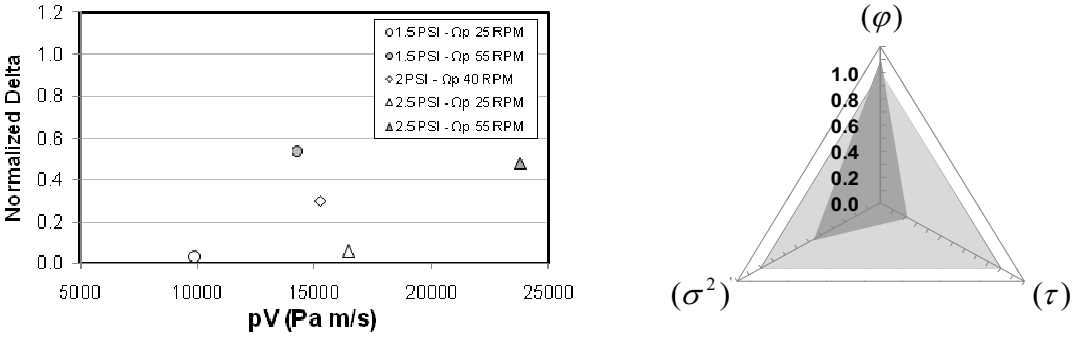
Figure 7(a) shows the removal rate as a function of the product of polishing pressure and sliding velocity (pV). The removal rate increases non-linearly with pV. However, unlike the trend of average and variance of shear force, the removal rate at polishing pressure of 2.5 PSI and platen rotational rate of 25 RPM lies in the expected range. Figure 7(b) shows the normalized RR WIWNU using the same reference process as shown in Figure 4(b).

Figure 8(a) shows the normalized 'delta' values using the same reference process as shown in Figure 5(a). The normalized 'delta' value at the polishing pressure of 2.5 PSI and platen rotational rate of 25 RPM was close to that at the polishing pressure of 1.5 PSI and platen rotational rate of 25 RPM. Such a low 'delta' value is due to its low shear force, low variance of shear force coupled with its higher removal rate for a shorter required polishing time. The normalized 'delamination triangle' is shown in Figure 8(b).



(a) (b)

Figure 7. (a) Removal rate and (b) normalized RR WIWNU as a function of pV



(a) (b)

Figure 8. (a) Normalized 'delta' values and (b) normalized 'delamination triangle' at the polishing pressure of 2.5 PSI, slurry flow rate of 200 ml/min and Ω_p/Ω_w of 25 RPM/148 RPM

Conclusion

In this study, a 'delamination triangle' consisting of average shear force, variance of shear force, and required polishing time is defined, and normalized 'delta' is calculated based on the product of the above three components. Results show that optimizing polishing parameters, such as polishing pressure, wafer and platen rotational rates, and slurry flow rate, can significantly reduce the 'delta' values and improve the removal rate within-wafer-non-uniformity.

References

1. http://aracainc.com/products/apd-800
2. G. Muldowney, J. Hendron and T. Crkvenac, *Proceeding of the 10th International Conference on Chemical-Mechanical Polish (CMP) Planarization for ULSI Multilevel Interconnection (CMP-MIC)*, p. 168 (2005)

ECS Transactions, 18 (1) 473-478 (2009)
10.1149/1.3096488 © The Electrochemical Society

Single Dispersion to Polish both Bulk Cu and Residual Cu along with Barrier Layers

S. V. S. B. Janjam[a], D. Roy[b], and S. V. Babu[c]

[a] Center for Advanced Materials Processing, Clarkson University, Potsdam, New York 13699, USA
[b] Department of Physics, Clarkson University, Potsdam, New York 13699, USA

Integration of Cu/Ta/low-k dielectric structures requires that chemical mechanical planarization (CMP) of Cu lines and barrier layers be performed at low down pressures (≤ 2 psi) to avoid damages to the underlying fragile dielectrics. At the same time, to avoid loss of throughput, the chemical activity of the slurry must be increased. We report an approach to slurry engineering that can help to achieve these goals in CMP of Cu/Ta/TaN over low-k dielectrics. These slurries employ silica abrasives in a single dispersion of oxalic and tartaric acids, with the objective of removing both Cu lines and Ta/TaN barrier layers simply by varying the operating slurry pH. In this approach, bulk Cu can be removed at high rates and residual Cu can be removed at lower rates (to minimize dishing) by adjusting the pH and the H_2O_2 content of the slurries. Ta/TaN layers also have good polish rates.

Introduction

As the feature size in semiconductor devices drops below 32 nm, new challenges appear for the technique of chemical mechanical planarization (CMP), which is commonly used in the processing of interconnect structures (1). These new structures contain porous, mechanically fragile low-k dielectrics, and avoiding damages to these materials during CMP of their overlying Cu lines and Ta (and TaN) barrier layers is a nontrivial task (2). Minimization of surface defects due to corrosion, scratches, dishing, and surface contamination is a critical requirement of the CMP process, which often is difficult to adequately meet for the new systems using traditional chemistries of material removal (3).

To minimize mechanical damages to the fragile low-k material, it is necessary to perform the planarization process at reduced down pressures (1-3). Moreover, during CMP, these dielectrics become prone to chemical damages that often are caused by alkaline slurries (4). Therefore, in addition to reducing the down-pressure of CMP, a careful design of the slurry chemistry, possibly avoiding alkaline media, is necessary to support efficient, chemically dominant CMP of Cu and Ta/TaN in interconnect structures containing low-k dielectrics.

In this work, we demonstrate that the aforementioned requirements of low-pressure polishing and non-alkaline CMP environment can be effectively met by using oxalic acid (OA) or tartaric acid (TA) as a surface-complexing agent, along with silica abrasives and H_2O_2 as an oxidizer. The present work is a continuation and extension of our single-slurry CMP study that was recently initiated using OA based slurries (5). The slurry chemistry of these systems supports a single-dispersion approach to CMP of both Cu lines and Ta/TaN barrier layers. Ta and Cu discs, as well as 2” diameter wafers were used for the polishing experiments reported in the present work. The results presented here

Ex. 1016, Page 501

demonstrate that the selectivity of Ta:Cu polish rates can be tuned through adjustments of the slurry pH. Of course, they still need to be validated for patterned structures.

Experimental Procedure

Cu disc planarization

A 99.99% pure Cu disc of 1" diameter was used as polishing sample for the experiments performed using a Struer's Labopol-5 polisher. The polish table (platen) was rotated at 90 rpm, while the disc was held stationary with an applied down-pressure of either 6.3 or 2 psi. Approximately 60 ml/min of the slurry was dispensed onto the center of the pad during the polishing process. The slurry contained in the beaker was continually stirred to maintain a good dispersion. Politex, X-Y groove polishing pads were used and each polishing experiment (of 2 min duration) was repeated three times. The reported polish rate is the average of the three such polish runs.

2" Wafer planarization

A semi-automated Center for Tribology Research (CETR) polisher was used to polish 2" blanket copper/tantalum/tantalum nitride films with a starting film thickness of ~ 3000 Å. The carrier and platen were rotated in the same direction and with the same speed (90 rpm) to obtain a constant linear velocity between the wafer and the pad. All the wafers were polished at a down-pressure of 2 psi. Slurries contained fumed silica (Aerosil 130) or colloidal silica (Nexsil 35A ~ 35 nm) abrasive particles and pH-adjusted solutions of H_2O_2 with TA or OA. Chemical solutions prepared with deionized water were dispensed onto the pad during polishing at a flow rate of ~ 120 ml/min. The slurry was stirred using a magnetic stirrer during polishing to maintain a good dispersion. A politex X-Y-grooved pad was used for these polishing experiments also.

Results and Discussion

Fig. 1 shows calculated pH dependent speciation diagrams for tartaric and oxalic acids, used in this work as surface complexing agents for CMP of Ta, TaN and Cu.

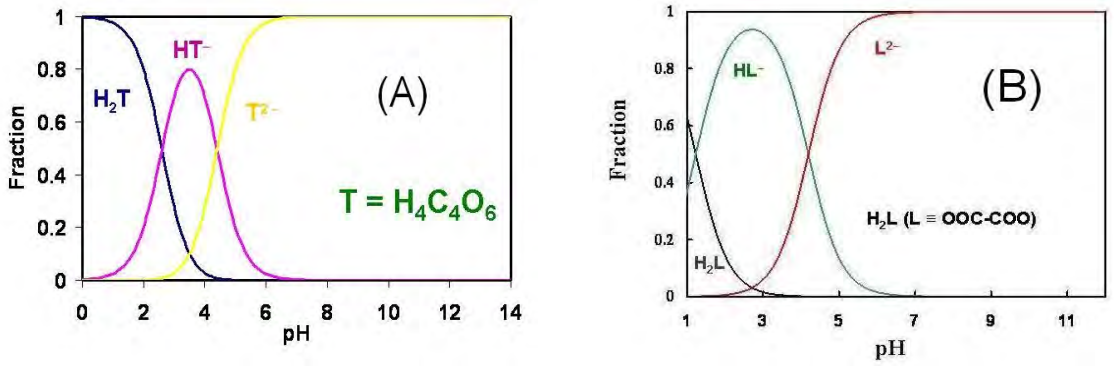


Figure 1. Calculated speciation diagrams for (A) tartaric and (B) oxalic acids.

The main non-alkaline region used here for the slurry development is between 3.0 and 6.0, with some exploratory measurements performed at pH = 8.0. This pH range utilizes the different surface chemistries of the HT^- and T^{2-} (or the HL^- and L^{2-}) species of TA (or

OA) adsorbed onto Ta and Cu. Detailed results of Cu (disc and wafer) polish rates (PRs) in the OA slurries have been discussed in our previous papers (6, 7) and are excluded from the present report. The PRs for Cu recorded with both these acids exhibited similar pH dependencies due to the similar chemistries of TA and OA on Cu. In all these experiments, considerable removal of Cu was observed in the acidic region, but these PRs dropped considerably at pH > 5.0 (6, 7).

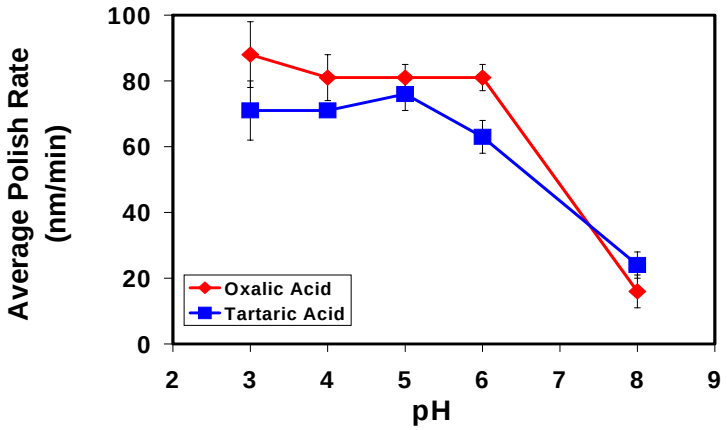


Figure 2. Ta disc polish rates as a function of pH at 6.3 psi for the slurry containing 3 wt. % fumed silica (Aerosil 130) + 5 wt. % H_2O_2 and (A) 0.13M OA (B) 0.13M TA.

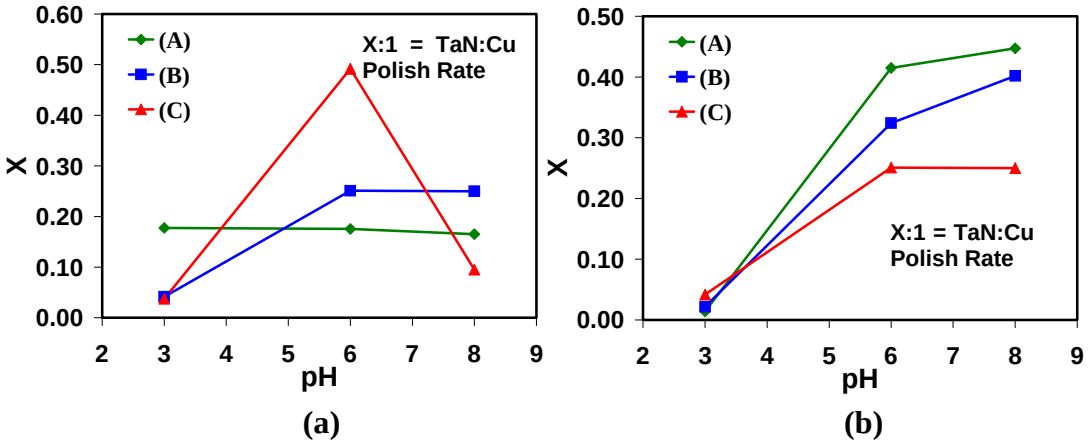


Figure 3. (a) TaN:Cu polish rate selectivity as a function of pH at 2 psi when the slurry contains 0.13M TA, 5 wt. % colloidal silica and (A) 1 wt. %, (B) 5 wt. % and (C) 10 wt. % H_2O_2. (b) TaN: Cu polish rate selectivity as a function of pH at 2 psi when the slurry contains , 5 wt. % colloidal silica, 5 wt. % H_2O_2 and (A) 0.032M, (B) 0.065M and (C) 0.13M TA.

Fig. 2 shows illustrative results for Ta polish rates in OA and TA based slurries. For both acids, appreciable (and nearly pH-independent) PRs of Ta were observed at pH = 3.0-6.0. The PR measurements shown in Fig. 2 were repeated for Ta and TaN wafers at a down pressure of 2.0 psi. The pH dependent PR results of these latter experiments also were similar to those shown in Fig. 2. In the acidic region Ta showed good removal

rates as Cu did under the same conditions. At $5.0 \leq pH \leq 7.0$, however, Cu-PRs dropped while Ta-PRs still remained nearly unchanged.

The Ta:Cu and TaN:Cu selectivity parameters were evaluated using the PR data for Ta, TaN and Cu blanket wafers, recorded in TA and OA slurries. Illustrative results of this selectivity analysis for TaN are presented in Fig. 3. A full set of PR results obtained for Ta, TaN and Cu using TA and OA based slurries is summarized in the first three columns of Table 1. The corresponding PR-selectivity parameters are listed in the last two columns of this Table.

Table I. Values of polish rates (PRs) and polish rate selectivity for TaN, Ta and Cu wafers polished at 2.0 psi using 5 wt% colloidal silica + 5 wt% H_2O_2 + 0.13M tartaric or oxalic acid at different pH settings.

Slurry	TaN PR (nm min^{-1})	Ta PR (nm min^{-1})	Cu PR (nm min^{-1})	TaN:Cu PR Selectivity	Ta:Cu PR Selectivity
5 wt% silica + 5 wt% H_2O_2 + 0.13 M TA at pH = 3.0	41 ± 4	42 ± 3	915 ± 19	0.05:1	0.05:1
5 wt% silica + 5 wt% H_2O_2 + 0.13 M TA at pH = 6.0	73 ± 5	60 ± 5	154 ± 9	0.47:1	0.39:1
5 wt% silica + 5 wt% H_2O_2 + 0.13 M OA at pH = 3.0	101 ± 5	65 ± 7	> 4000	<0.03:1	<0.02:1
5 wt% silica + 5 wt% H_2O_2 + 0.13 M OA at pH = 6.0	105 ± 9	60 ± 3	360 ± 9	0.29:1	0.17:1

An examination of the selectivity parameters presented in Table 1 indicates how the TA and OA slurry chemistries could be useful for preferential removals of Ta/TaN and Cu. Based on the results of Table 1, bulk Cu can be removed by polishing in the acidic region of the OA/TA slurries. Subsequently, any residual Cu as well as Ta/TaN can be polished in the weakly acidic and neutral pH regions. Thus, mere adjustments of the slurry-pH control selective removals of Ta and Cu at different stages of planarization. This selectivity of material removal using the single dispersion slurry (single complexing chemistry for both Cu and Ta/TaN) is the main finding of the present study. The Cu removal rates can be significantly suppressed forther by judicious use of surfactants like Ammonium dodecyl sulfate and Dodecyl benzene sulfonic acid (8, 9).

The surface chemistries of the OA and TA complexing agents have been tested using electrochemical techniques. Based on the results of these measurements, the following (chemically dominant) mechanism of Cu/Cu-oxide removal in these systems is proposed. In both OA and TA systems, Cu_2O and CuO are formed by H_2O_2 on Cu. In the

acidic environment, Cu dissolves as follows: $Cu = Cu^{2+} + 2e^-$; $Cu + H_2O_2 + 2H^+ = Cu^{2+} + 2H_2O$. In OA ($H_2L$, $L \equiv C_2O_4$), CuO reacts with HL^- to form soluble $[CuL_2]^{2-}$. Cu can also react with HL^- (pH <5) to form soft films of CuL, that in turn can be removed by low-pressure (≤ 2psi) polishing. In TA (H_2T, $T \equiv C_4H_4O_6$), Cu_2O and CuO can react with HT^- (pH ≈ 2-5) and T^{2-} (pH >4), respectively, to form soluble $(CuT_2)^{2-}$. Mechanical abrasion removes these species buried within the Cu-oxide layers.

The following (chemically dominant) mechanism of Ta/Ta-oxide removal is proposed: Ta_2O_5 (native oxide film on Ta), combined with catalytic breakdown of H_2O_2 leads to surface sites of $Ta(OH)_5$ in both OA and TA. In OA, $Ta(OH)_5$ reacts with HL^- (pH < 5) as well as with L^{2-} (pH > 3.5) to form the soluble complex $[TaO_4L_2]^{3-}$. In TA, $Ta(OH)_5$ reacts with HT^- (pH ≈ 2-5) and T^{2-} (pH >4) to form the soluble complex $[Ta(O_2)_3T]^{3-}$. These species are buried in porous Ta_2O_5, weakening the structural integrity of the oxide film, and the latter can be removed by mechanical abrasion at low pressure.

The Ta sites on TaN are likely to be oxidized as Ta_2O_5 in the aqueous (oxygen and H_2O_2 containing) CMP environment. Thus, removal of TaN in the presently used slurries also follows essentially the same chemical route of Ta removal mentioned above. In other words, CMP of Ta and TaN is chemically supported by the generation of structurally fragile films of surface complexes. The actual removal of these complexes occurs through (low-pressure) mechanical abrasion of the porous layers of the surface complexes and oxides.

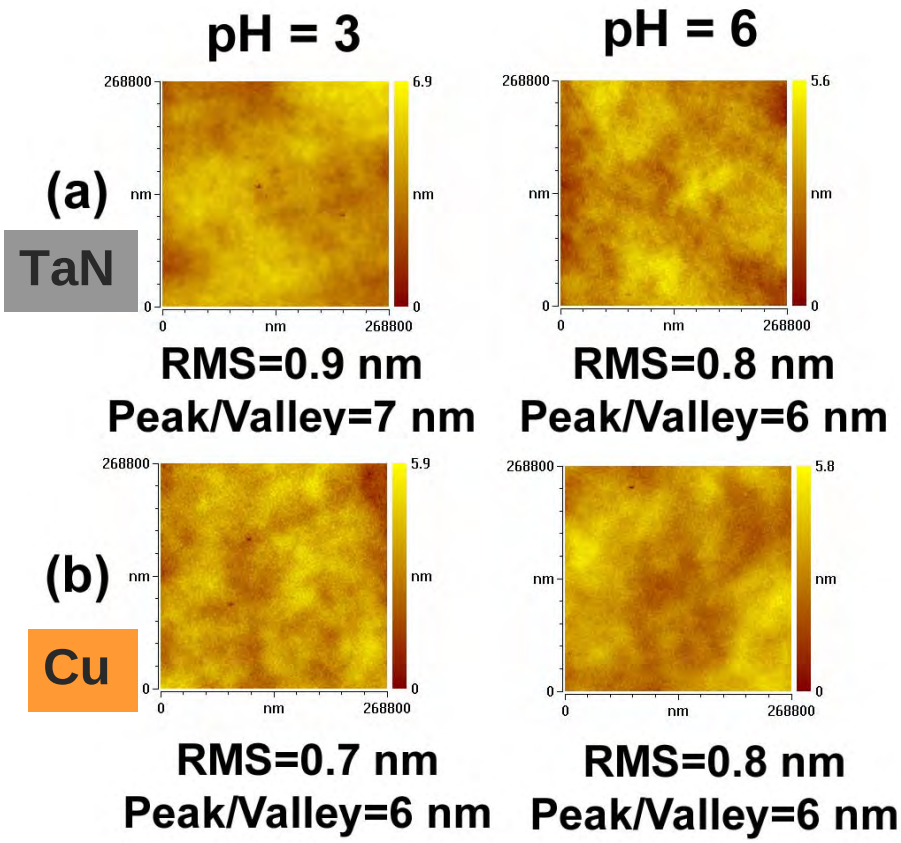


Figure 4. Surface roughness profiles of (a) TaN- and (b) Cu- wafers after CMP using slurries containing 5 wt. % colloidal silica, 0.13M TA and 5 wt. % H_2O_2 at pH = 3 and 6.

Optical profilometry was used to study the surface morphologies of Cu, Ta and TaN wafer films polished using the OA and TA based slurries. Some profilometry data obtained from these measurements for Ta wafers have been shown in a recent publication (6). Illustrative results for TaN and Cu wafers are presented here in Fig. 4. The surface quality of each polished sample was evaluated in terms of the peak-to-valley (P/V) and root mean square (RMS) roughness parameters (noted in Fig. 4). For all the (Ta, TaN and Cu) samples tested in this way, the post-polish surface roughness always was found to be $\leq$ 1 nm, and the corresponding P/V distances were $\leq$ 10 nm. These values were comparable to those of as-received wafers, and hence, indicated good defect-free morphologies of the wafers that were CMP-treated using the TA and OA based slurries.

Conclusions

The experimental results presented in this work show that pH-adjusted slurry solutions of OA or TA with H_2O_2 can be useful to achieve selective planarization of both Cu lines and Ta/TaN barriers in CMP. A rather broad pH-range (3.0-7.0) is available to perform these planarization steps. Bulk Cu removal can be performed near the lower limit of the pH range, and residual Cu can be cleared at the upper end of this range. At the same time, Ta/TaN removal should be appreciable and essentially unchanged within these limits of pH variations. Since this pH-range is non-alkaline, the slurries reported here should be advantageous for avoiding/minimizing OH^- induced chemical damages (carbon depletion, dangling bond formation) that many Si-based low-k dielectric layers (underlying the Ta/TaN barrier) are susceptible to.

Acknowledgments

This work was supported in part by IBM through the Semiconductor Research Corporation and by ARO (W911NF-05-1-0339).

References

1. International Technology Roadmap for Semiconductors, ITRS (2006).
2. E. Matijevic and S.V. Babu, *J. Colloid Inter. Sci*. **320**, 219 (2008).
3. M. R. Oliver, Ed. “Chemical Mechanical Planarization of Semiconductor Materials”, *Springer*, New York, 2006.
4. T.C. Chang , Y.S. Mor, P.T. Liu, T.M. Tsai , C.W. Chen, Y.J. Mei and S.M. Sze, *Thin Solid Films*, **398**, 523 (2001).
5. S. V. S. B. Janjam**,** V. V. S. C. Surisetty, S. Pandija, D. Roy and S. V. Babu, *Proc. ISTC Conf. 2008*, The Electrochemical Society, Pennington (2008), p. 561.
6. S. V. S. B. Janjam, C. V. V. S. Surisetty, S. Pandija, D. Roy, and S. V. Babu, *Electrochem. Solid-State Lett.* **11**, H66 (2008).
7. S. V. S. B. Janjam, S. Peddeti, D. Roy and S. V. Babu, *Electrochem. Solid-State Lett.* **11,** H327 (2008).
8. Y. Hong, U. B. Patri, S. Ramakrishnan, D. Roy and S. V. Babu, *J. Mater. Res.* **20,** 3413 (2005).
9. C. V. V. S. Surisetty, P. C. Goonetilleke, D. Roy and S. V. Babu, *J. Electrochem. Soc.* **55**, H971 (2008).

ECS Transactions, 18 (1) 479-484 (2009)
10.1149/1.3096489 © The Electrochemical Society

Ultra-High Removal Rate Copper CMP Slurry Development for 3D Application

B. Hu[a], H. Kim[a], R. Wen[a] and D. Mahulikar[b]

a Planar Solutions LLC, 6550 South Mountain Road, Mesa, AZ 85212, USA
b Planar Solutions LLC, 80 Circuit Dr. N. Kingstown, RI 02852, USA

Chemical-mechanical planarization (CMP) has been widely accepted in IC manufacturing. With technology moves to 45 nm and beyond, the challenge of continuous reduction of interconnect features resulting in the increased Cu line resistivity. To mitigate this effect 3-D packaging where through-silicon vias (TSV) are filled with electroplated Cu can be used. A moderately high removal rate slurry, ER9200 was developed and the work was presented elsewhere (ICPT Taiwan Conference Nov. 2008). However certain 3D structures have very thick Cu overburden. For such structures ultra high removal rates of 4~5 µm/min may be needed. This process requires a high speed Cu CMP process to remove the thick overburden Cu layer. Presented herein is a Cu CMP slurry with ultra-high removal rate, yet good Planarization Efficiency and low defects. This new colloidal silica based developmental slurry achieves much higher removal rate, and is suitable for Cu 3D CMP application.

Introduction

High density IC devices involve the stacking of multiple layers of interconnects. The process of chemical-mechanical planarization (CMP) provides a flat and smooth surface on the wafer for lithography needs and has become a critical process to achieve high yields in semiconductor manufacturing[1-2]. As the semiconductor industry moves toward finer structures (45nm and beyond nodes), the integration schemes are facing a challenge to balance the interconnect features reduction and limitation of the physical properties of the interconnect materials. Wafer-level monolithic processing has the performance advantage of reduced interconnect parasitic and lower interconnectivity cost. Through silicon via (TSV) stack die configurations by creating interconnect structure through wafer, which can reduce Joule heating issue and chip footprint while increasing interconnect density. It allows increased input–output counts and minimize the chip real estate[3].

TSV are electrical pathways formed vertically through the whole wafer thickness which can decrease the complexity of interconnect pathways and allow more input–output channels in the package. Copper can be used in the fabrication of TSVs as the interconnect material. In this Cu 3D TSV process, the bulk Cu deposited fills the vias and also deposits a thick, non-uniform copper layer across the wafer. The 3-20 µm over burden Cu layer need to be removed using CMP, leaving only the Cu in the vias creating a through-wafer interconnect (Fig. 1). It requires a novel and optimized Cu CMP process for removing this thick Cu layer[4]. One of major requirement is ultra-high speed Cu CMP process to achieve a reasonable process time and throughput.

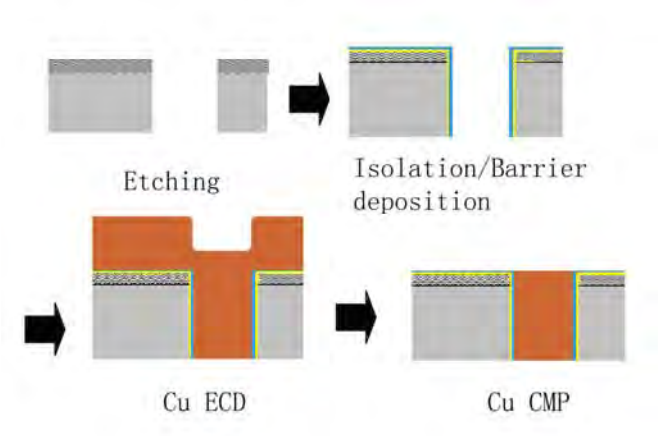


Figure 1. A Typical Cu TSV formation process

Generally, a Cu CMP slurry contains abrasive particles, oxidizer, corrosion inhibitor and pH adjuster. Optionally, complex agents, surfactants, rheological tuner, and other exotic ingredients are added in to achieve certain performance refer to US Patent 6083840[5]. CMP requires the balance of removal rates, planarization and surface quality. In order to achieve this balance, a corrosion inhibitor is introduced into the formulation. Its role is to passivate the surface so that the oxidation rate can be controlled, as well as to prevent pitting and etching [6]. The Cu removal rate is dependent on the surface oxidization rate, passivation rate, Cu etching rate, Cu dissolving rate and the polishing efficiency of the passivate film by the abrasive particle. High Cu removal rate may cause high surface corrosion, defectivity and bad topography control. The real challenge for developing 3D TSV Cu CMP slurry is how to achieve ultra-high Cu removal rate with good planarization, topography, defectivity and surface roughness.

Based on Planar's successful CSL-904X Cu slurry platform, we have developed new high removal rate Cu slurry and the work was presented at CPT Taiwan 2008[7]. For the further improvement, an ultra-high removal rate slurry, ER9212, was developed with target 4~5 μm/min removal rate for Cu 3D TSV application. It can also be used in Cu MEMS process. Slurry development for the 3D application is described herein.

Experiments

All experiments were performed on a Mirra Polisher (Applied Materials, Sunnyvale, CA) with IC1010 pads which were conditioned with either an ABT, 3M or Kinik conditioner. 200mm Cu Blanket wafer, Cu slug and MIT 854 pattern wafers have been used in this study. The removal rate was determined by measuring the weight loss and Cu thickness by sheet resistance on Resmap with a 4-point probe. Corrosion response was measured using electro-chemical techniques[8]; the defectivity was measured on AIT-UIV. Topography was measured using Veeco AFM profiler. During CMP operations, many process variables must be taken into consideration to obtain a robust and stable process. After optimization, the machine variables such as down force, table speed, head speed and conditioning were held constant throughout the experiment for data collection.

The Planar Solutions experimental slurry ER9212 was selected in this study. ER9212 is developed based on Planar's successful CSL-904X Cu slurry platform. It has higher Cu

Ex. 1016, Page 508

bulk removal rate with good uniformity and defectivity. It is designed for Cu CMP applications needing high removal rate, such as Cu 3D TSV, Cu MEMS and Top layer Cu planarization. Table below shows the relative properties of this slurry.

TABLE 1. The slurry properties of ER9212

Product Name	**ER9212 Column Header Goes Here**
Abrasive particle	Colloidal Silica
Oxidant	Hydrogen Peroxide
pH	Neutral pH
Total Heavy metal	<10 ppm
Typical Cu removal rate	>4 µm/min at 3~4psi down force
Cu Corrosion Rate	<10 Å /min

Result and Discussion

We initially looked at acidic platform slurries. Because of the fast Cu oxidization rate and dissolution rate in the acidic condition, these slurries can polish at very high removal rates. In our screen test, most acidic slurries showed high removal rates (Fig 2). But most of the polished wafers showed surface corrosion issue. A pad stain was also observed during the CMP process, which will impact the process stability, reduce the pad life and increase the defectivity. Therefore another approach was considered.

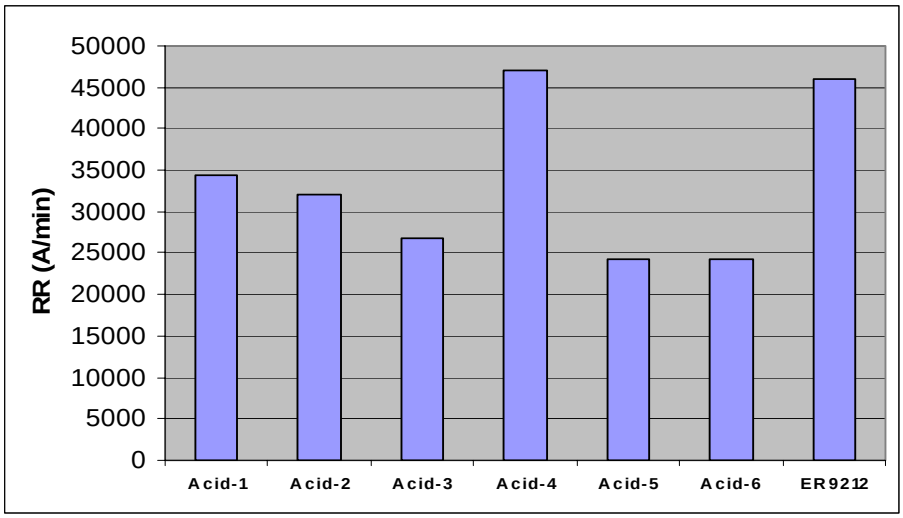


Figure 2. The chemistry screen for acidic slurry

Planar Solutions LLC has successfully commercialized a series of slurries based on a neutral pH CSL904X platform. Typical Cu bulk removal rates of these slurries range from 8000 to 11000 Å /min. These rates are not adequate for the 3D application. These slurries have a Cu removal rate enhancer and a Corrosion Inhibitor. As discussed in the ICPT Taiwan paper, if we simply increased the removal rates enhancer content the rates would go up but so would topography and surface roughness (Fig 3). Our first attempt was with a high removal enhancer. While a removal rate of 3~5 µm/min can be achieved, all the Cu in the trenches was dissolved during the pattern wafer polishing. The surface roughness after Cu CMP was around 25-30 Å, which is unacceptable. It indicated high corrosion generated during CMP. By optimizing the slurry chemistry, we obtained a new slurry which yielded optimum performance. This was called ER9212. The performance showed this composition may be suitable for 3D TSV Cu CMP process.

Ex. 1016, Page 509

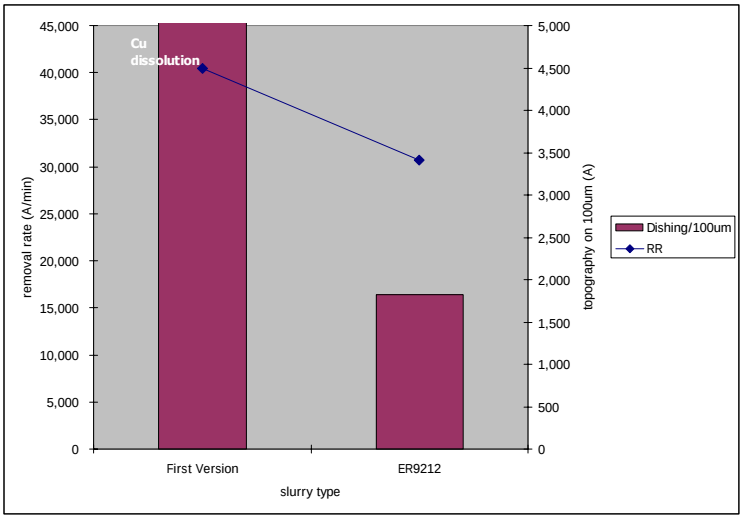


Figure 3. The chemistry impact on removal rate and topagraphy

Since much thicker copper layer needs to be removed in Cu 3D TSV, higher and tunable RR is required for this application. We introduced the Factor A to control the removal of ER9212. The removal rate was obtained under 4psi polishing down force as seen in Figure 4. The removal rate of ER9212 is tunable by Factor A. We can achieve removal from 2.5µm/min to 7.7µm/min. It will give customers an option to choose the removal rate they need based on their Cu TSV integration and CMP process requirement.

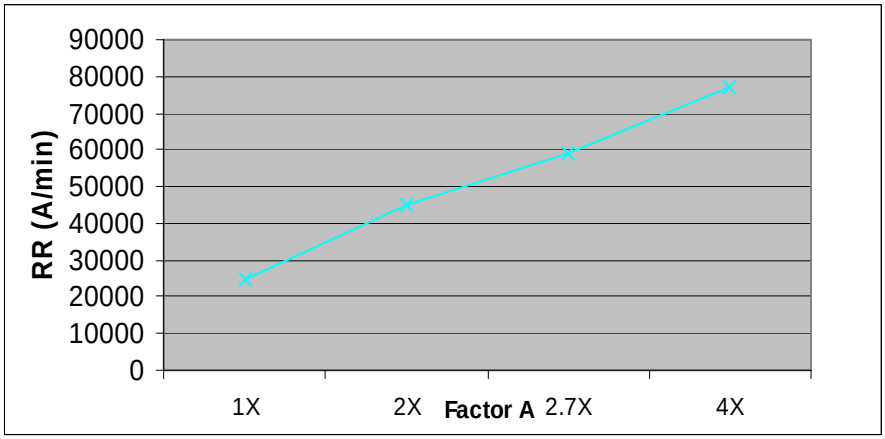


Figure 4. The tunable removal rate ER9212

The removal rate and polishing pad temperature were collected under different polishing down force (Fig. 5). The result showed that the removal rate increases with polishing down force. The higher polish down force resulted higher removal rate. The Cu bulk removal rate of ER9212 was above 4µm/min at a down force above 3 psi. The temperature on the pad is also a key parameter for high down force Cu CMP since too high temperature (>75˚C) will cause the pad delamination and CMP tool failure issues. The mechanical friction of CMP and heat generated by chemical reaction are the major factors infect the pad temperature. Higher pad temperature was seen with increasing polishing down force. Reduced rotation speed and increased slurry flow rate will help to

Ex. 1016, Page 510

reduce the pad temperature during the CMP process. The polishing pad temperature of ER9212 is around 70˚C even at 5psi down force. The pad temperature is expected to be a lower with real patterned wafers.

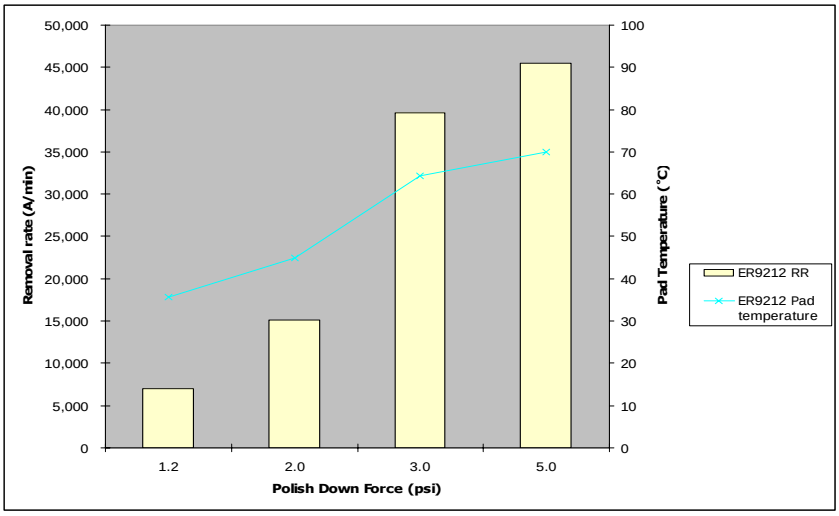


Figure 5. The removal rate and pad temperature vs polishing down force

For the high speed Cu CMP process, very high concentration by-products will be seen on the pad. These Cu by-products will stain the pad at high down force and high temperature, which will reduce the Cu removal rate and increase the defectivity. The drop of the Cu removal rate will also impact the CMP process control, such as end point and over polishing control. No pad stain was observed during the polishing process for ER9212. After polish under 2psi down force, about 1.5 µm of Cu was removed and less than 10 Å Ta was removed at the same condition. The selectivity of the Cu/Ta is greater than 1500:1, which indicates that ER9212 is high selective Cu slurry.

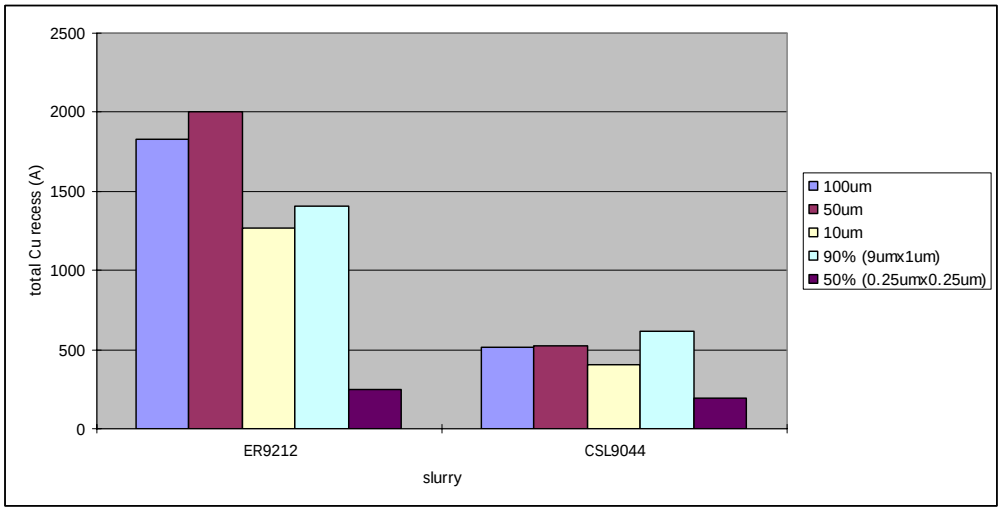


Figure 6. Topography of MIT 854 pattern wafer after polished with ER9212

MIT 854-TEOS pattern wafer were used for the topography study (Fig. 6). The over burden Cu layer on this wafer is around 1.0 µm. The wafer was polished under 1.2psi

down force till end point. The polished wafers showed no Cu residue on the wafer. The dishing on 10 µm line of high removal rate slurry is around 1265 Å, which is acceptable for Cu 3D TSV application. Since the major bottle neck of the Cu CMP for 3D TSV application is the bulk polishing step, much lower topography can be achieved if a low dishing slurry, such as Planar Solutions CSL9044C, was used at the soft landing step.

The surface roughness was measured on the Cu blanket wafers after 0.6 µm Cu layer was removed. ER9212 is neutral pH at point of use with low Cu surface corrosion. After polishing, good nonuniformity was observed. The finish Cu wafer surface was very smooth and showed low surface roughness, which between 9Å -13Å. The result of electro-chemical experiments also showed the 9.3Å /min corrosion rate of ER9212 on the Cu wafer surface. It shows this slurry has low corrosion and high planarization efficiency.

The result of this study indicated that the Cu complexing and Cu surface passivation is very important in high-speed Cu CMP slurry design. Well tuning the balance of these factors gives high Cu removal rate and maintains the decent topography, surface roughness and good defectivity, which is important for 3D TSV Cu CMP application.

Conclusion

Planar's ER9212, a developmental composition, gives high Cu bulk removal rate, decent topography, and low surface roughness. It is a good slurry candidate for high speed Cu CMP applications, such as Cu 3D TSV, Cu MEMS and top layer Cu CMP.

Acknowledgements

The authors acknowledge the technical help by M. Tanaka, L. Gramm, G. Moyaerts and K. Cline. We thank Planar Solutions LLC for permission to publish this paper.

References

1 J. M. Steigerwald, S. P. Murarka, and R. J. Gutman, *Chemical Mechanical Planarization of Microelectronic Materials*, p. 15, John Wiley & Sons, New York (1997).

2 T. Tucker, *Chemical-Mechanical Planarization of Semiconductor Materials*, M. R. Oliver (Ed.), p. 137, Springer-Verlag, Berlin (2004).

3 T. Tessler, in *3rd IC Packaging Technology Exposition and Conference*, Japan (2002).

4 S. L. Burkett, X. Qiao, D. Temple, B. Stoner, and G. McGuire, J. Vac. Sci. Technol.B, 22, 248 (2004).

5 B. Mravic, D. Mahulikar, US Patent 6083840A

6 D. Mahulikar, R. Wen and B. Hu, Specialty Chemicals Magazine, March, 25(3), 2 (2005).

7 B. Hu, R. Wen, H. Kim and D. Mahulikar, "High Removal Rate Copper CMP Slurry Development"; ICPT 2008, Taiwan, (2008)

8 R. Wen, et al.; Lake Placid CMP conference proceedings, Lake Placid, NY (2005).

ECS Transactions, 18 (1) 485-490 (2009)
10.1149/1.3096490 © The Electrochemical Society

Effect of H_2O_2 Concentration and Slurry Dilution on Removal Rate and Surface Quality for a Cu Slurry

Jie Lin and Tianbao Du

Fujimi Corporation, 11200 SW Leveton Drive, Tualatin, Oregon 97062, USA

The effects of H_2O_2 concentration and slurry dilution on Cu removal rate and surface quality were investigated for a colloidal silica-based slurry. At 1:6 dilution, the Cu removal rate increased with increasing H_2O_2 concentration, reaching a maximum at 0.90 wt% (H_2O_2 threshold concentration), and then decreasing and leveling off at higher H_2O_2 concentrations. A smooth Cu surface was obtained above the H_2O_2 threshold, but a rough Cu surface was produced below the H_2O_2 threshold. At higher dilutions of 1:8 and 1:10, similar effects of H_2O_2 concentration on Cu removal rates and surface qualities were observed, but the H_2O_2 threshold shifted down to 0.74 and 0.66 %, respectively. Electrochemical measurements explain the observation of a rough Cu surface below the H_2O_2 threshold, and a smooth Cu surface above the H_2O_2 threshold.

Introduction

Chemical mechanical polishing (CMP) has been widely used in manufacturing IC devices containing copper multilevel metallization. In Cu CMP, a high removal rate Cu bulk slurry can be used to remove the overburden copper and planarize the wafer surface quickly and cost effectively. The most common oxidizer for use in Cu CMP is H_2O_2. Selection of H_2O_2 concentration is important in order to achieve high Cu removal rate and good Cu surface quality. Several groups reported the effect of H_2O_2 concentration on Cu removal rate [1-5]. However, the effect of H_2O_2 concentration on Cu surface quality has not been reported. In addition, the impact of slurry dilution on the H_2O_2 threshold concentration and the resulting Cu surface quality has not been investigated.

In this paper, interactive effects of H_2O_2 concentration and slurry dilution on the Cu removal rate and Cu surface quality is reported and discussed for a Cu bulk slurry (FCC1024M). A mechanism is proposed based on electrochemical measurements using a potentiodynamic polarization technique under in-situ and ex-situ conditions.

Experimental

FCC1024M is a colloidal silica slurry that contains a Cu complexing agent, a Cu corrosion inhibitor, and uses H_2O_2 as an oxidizer. Working slurry was prepared by dilution of the slurry with DI water and addition of various amounts of 30 wt% H_2O_2. Polishing tests were performed with 200 mm Cu blanket test wafers on an Ebara EPO222 polisher using an IC1000 K-groove/Suba IV pad. Typical polishing process conditions were as follows: platen/head speed, 95/90 rpm; pressure, 3.5 psi; slurry flow rate, 260 mL/min; and polishing time, 45 sec. Ex-situ pad conditioning was done using a 100-grit diamond disk with down force of 125 Newton, platen/conditioner speed of 20/12 rpm, and conditioning time of 17 sec. Cu removal rate was determined by four-point probe

Ex. 1016, Page 513

resistivity measurement on a CDE ResMap. Cu surface quality was evaluated on an Olympus MX50 optical microscope. Cu surface roughness (R_a) was measured on a Tencor P-12 Profiler.

Potentiodynamic polarization tests were performed using an EG&G Princeton Applied Research model 2273 potentiostat/galvanostat with a three-electrode system. A Cu disc with cross-section area of 38 cm^2 was used as the working electrode, a stainless steel plate as the counter electrode, and a saturated Ag/AgCl electrode as the reference electrode. Potentiodynamic polarization sweep of the working electrode was performed from -0.5 to +0.5 V (vs. OCP, i.e. open circuit potential) at a sweep rate of 5 mV/s. In-situ electrochemical measurements during polishing were performed using a modified Strasbaugh polisher. A slurry build up of at least 10 mm was maintained above the pad in order to have a proper electrical contact between the working electrode and reference electrode. The platen was rotated at 33 rpm and the applied down force was 1 psi.

Results and Discussions

Effect of H_2O_2 concentration on Cu removal rate and surface quality

Figure 1 shows Cu removal rate vs. H_2O_2 concentration of FCC1024M at 1:6 dilution. The Cu removal rate increased with H_2O_2 concentration, reaching a maximum at 0.90 wt% (H_2O_2 threshold concentration) [4], and then decreasing and leveling off as more H_2O_2 is present. The effect of H_2O_2 concentration on Cu removal rate observed in this study is consistent with previous reports [1-5].

The impact of slurry dilution with DI water on the effect of H_2O_2 concentration was also investigated in this study. Figure 2 shows that, at higher dilutions of 1:8 and 1:10, similar effects of H_2O_2 concentration on the Cu removal rate were observed. However, the H_2O_2 threshold shifted down to 0.74 and 0.66 %, respectively.

The effect of H_2O_2 concentration on the Cu surface quality is shown in Figure 3 for 1:6 dilution. A rough Cu surface was produced when the H_2O_2 concentration was lower than the H_2O_2 threshold. The roughest Cu surface was observed at 0.66 % H_2O_2 concentration, below the H_2O_2 threshold. Only when the H_2O_2 concentration was significantly higher than the threshold (in the range where the Cu removal rate leveled off), a smooth Cu surface was obtained. Similar effects were observed for this slurry at 1:8 and 1:10 dilutions. Figure 4 presents the Cu surface roughness (R_a) vs. H_2O_2 concentration at 1:10 dilution. It shows that the roughness first increased with H_2O_2 concentration, reaching a maximum before the Cu removal rate peak (i.e. H_2O_2 threshold) is reached, and then decreasing and leveling off at higher H_2O_2 levels. A H_2O_2 concentration higher than 1.0 % is needed for a smooth Cu surface at 1:10 dilution.

Electrochemical measurements

Potentiodynamic polarization curves for Cu in the slurry at 1:10 dilution containing different H_2O_2 concentrations are presented in Figure 5A. Copper does not show passive metal behavior, and no obvious passivation region can be observed in these curves. Figure 5B shows corrosion potential and current vs. H_2O_2 concentration. The corrosion potential became more positive with increasing H_2O_2 concentration and leveled off above the H_2O_2 threshold. This could be due to the inhibited anodic reaction and the enhanced cathodic reaction in the presence of H_2O_2 [5]. The corrosion current displays a similar trend to that of Cu removal rate vs. H_2O_2 concentration, indicating that electrochemical dissolution of Cu plays an important role in the removal of Cu during polishing [3,6,7].

Ex. 1016, Page 514

The in-situ potentiodynamic polarization curves for Cu in the slurry at 1:10 dilution containing two different H_2O_2 concentrations are presented in Figure 6. At 0.3 % H_2O_2 concentration, the in-situ potentiodynamic polarization measurement during polishing shows much larger anodic current and lower corrosion potential than the ex-situ measurement without polishing. This indicates that the protective film formed on the Cu surface can be removed during polishing, resulting the exposure of a fresh Cu surface. However, at 1.5 % H_2O_2 concentration, the in-situ potentiodynamic polarization curve shows no obvious difference compared to the ex-situ curve, which indicates that the protective film formation rate is faster than its removal rate. Thus, a protective film is always present on Cu surface and the removal rate is limited by the polishing of the protective film.

Mechanism and application

In general, oxidation of Cu by H_2O_2, together with complexation of Cu by inhibitor, forms a protective film on the Cu surface during CMP process. The Cu removal rate is controlled by relative rates of the formation and removal of the protective film. When the slurry concentration is constant (hence the concentration of inhibitor is constant) at a given dilution, the H_2O_2 concentration will determine the protective film formation rate, thus controlling the Cu removal rate.

- Below the H_2O_2 threshold, the Cu oxidation rate is slower than the polishing rate of the protective film. Thus, the Cu removal rate is limited by the Cu oxidation rate and increases with H_2O_2 concentration. Complete removal of the protective film will expose Cu metal to chemical etch and mechanical abrasion, which results in a rough Cu surface.
  - The roughest Cu surface is produced when the H_2O_2 concentration is high enough to quickly oxidize the Cu surface but not high enough to form a uniform protective film, which results in enhanced/localized corrosion.
- At the H_2O_2 threshold, a maximum Cu removal rate is achieved when the protective film formation rate is equal to its polishing rate. Therefore, the protective film will be polished off completely. The fresh Cu metal can be exposed, and a Cu rough surface is the result.
  - Cu surface is still rough at the H_2O_2 concentration slightly above the threshold value. This can be explained by variations in the formation and polishing rates of the protective film. The variations may sometimes cause the protective film polishing rate to be equal to or higher than its formation rate (locally or globally), which results in exposure of Cu metal and thus a rough Cu surface.
- Above the H_2O_2 threshold, Cu oxidation rate becomes faster than the protective film polishing rate. Thus, the Cu removal rate is limited by the protective film polishing rate. In this H_2O_2 concentration range, the protective film is always present during the polishing, which prevents Cu metal exposure and results in a smooth Cu surface.

It was observed that the H_2O_2 threshold concentration decreased from 0.90 % at 1:6 dilution, to 0.74 % at 1:8 dilution, and to 0.66 % at 1:10 dilution. This decrease can be explained by the competition between the oxidation of Cu by H_2O_2 and the removal of the protective film by polishing. At a higher dilution, slurry component concentrations are lower, the protective film polishing rate is slower; thus, a lower H_2O_2 concentration is required for the oxidation of Cu to compete with the removal of the protective film to reach a balance between these two competing processes.

This study shows that, only when the H_2O_2 concentration in the slurry is significantly higher than the H_2O_2 threshold value, a smooth Cu surface can be obtained. Because slurry dilution will cause a shift in the H_2O_2 threshold, it is very important to choose a H_2O_2 concentration that is high enough to ensure a smooth Cu surface at all dilutions.

Conclusions

It is demonstrated in this study for a colloidal silica-based slurry (FCC1024M) that H_2O_2 concentration has effects not only on the Cu removal rate but also on the Cu surface quality. The effects result from a competition between the formation of a protective film on the Cu surface through Cu oxidation by H_2O_2 (as well as Cu complexation by the inhibitor) and the removal of the protective film by polishing. There is a H_2O_2 threshold concentration at which the Cu removal rate reaches a maximum. The H_2O_2 concentration must be significantly higher than the threshold value for a smooth Cu surface to be obtained. The H_2O_2 threshold concentration shifts down at higher dilutions, making it important to choose a H_2O_2 concentration that is high enough to ensure a smooth Cu surface at all dilutions.

Acknowledgments

The authors would like to thank the colleagues in the R&D Department at Fujimi Corporation, particularly Dr. Scott W. Rader, for assistance and discussion.

References

1. M. Hariharaputhiran, S. Ramarajan, S.V. Babu, in *Chemical-Mechanical Polishing Fundamentals and Challenges,* S.V. Babu, S. Danyluk, M. Krishnan, and M. Tsujimura, Editors, PV56, p. 129, The MRS Symposium Proceedings, Warrendale, PA (1999).
2. J. Hernandez, P. Wrschka, and G.S. Oehrlein, J. *Electrochem.Soc.*, **148**, G389 (2001).
3. S. Aksu, L. Wang, and F. M. Doyle, *J. Electrochem. Soc.*, **150**, G718 (2003).
4. F. M. Doyle, L. Wang, in *Proceedings of the 12th International VLSI Multilevel Interconnection Conference*, p. 267 (2003).
5. T. Du, D. Tamboli, V. Desai, and S. Seal, *J. Electrochem. Soc.*, **151**, G230 (2004).
6. P. Wrschka, J. Herbabdez, G. S. Oehrlein and J. King, *J. Electrochem. Soc.,* **147**, 706 (2000).
7. T. Du, D. Tambolic, and V. Desai, *Microelectron. Eng*. **69**, 1 (2003).

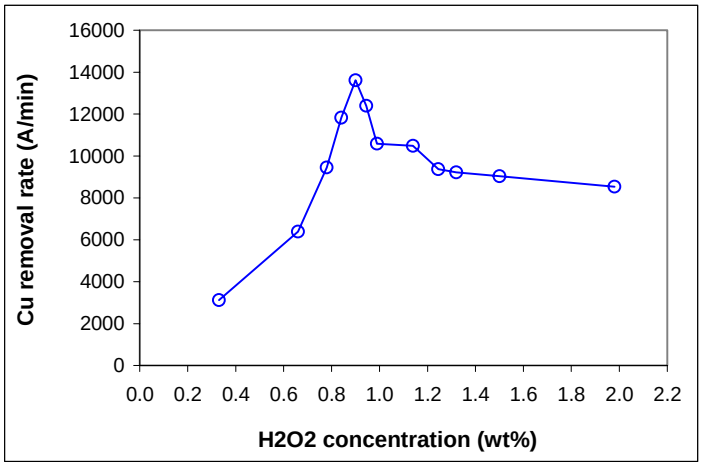


Figure 1. Cu removal rate vs. H_2O_2 concentration for FCC1024M at 1:6 dilution.

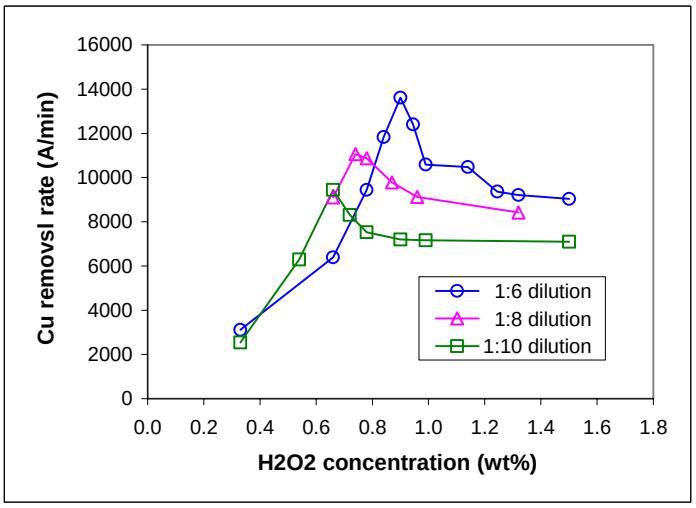


Figure 2. Effect of H_2O_2 concentration on Cu removal rate for FCC1024M at different dilutions.

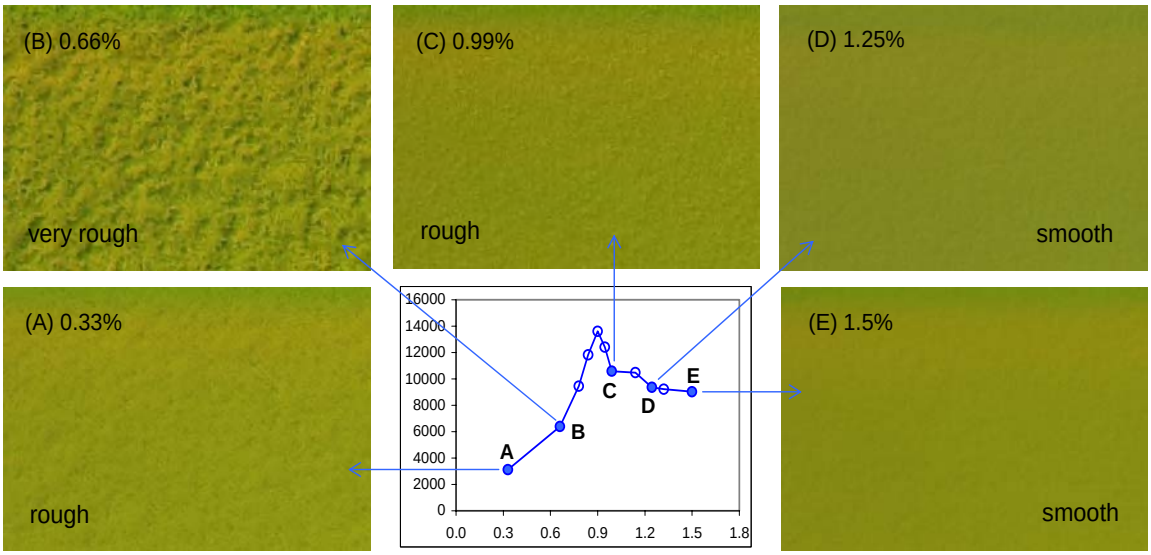


Figure 3. Microscope pictures of Cu surface after polishing with FCC1024M at 1:6 dilution containing different H_2O_2 concentrations.

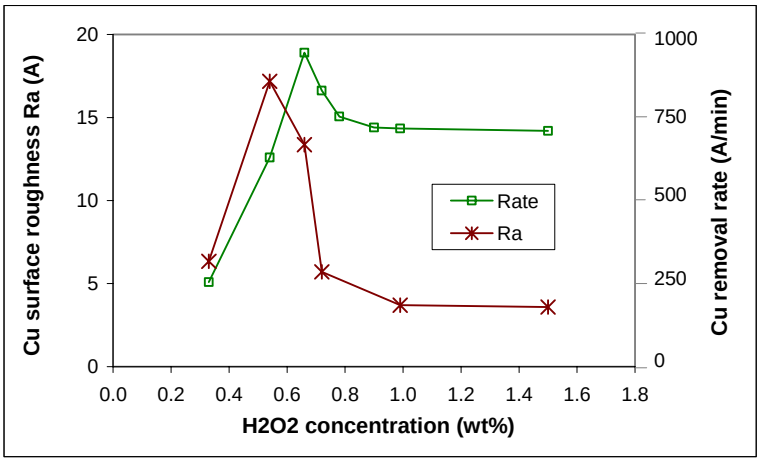


Figure 4. Effect of H_2O_2 concentration on Cu surface roughness (R_a) for FCC1024M at 1:10 dilution. Cu removal rate as a function of H_2O_2 concentration is shown for comparison.

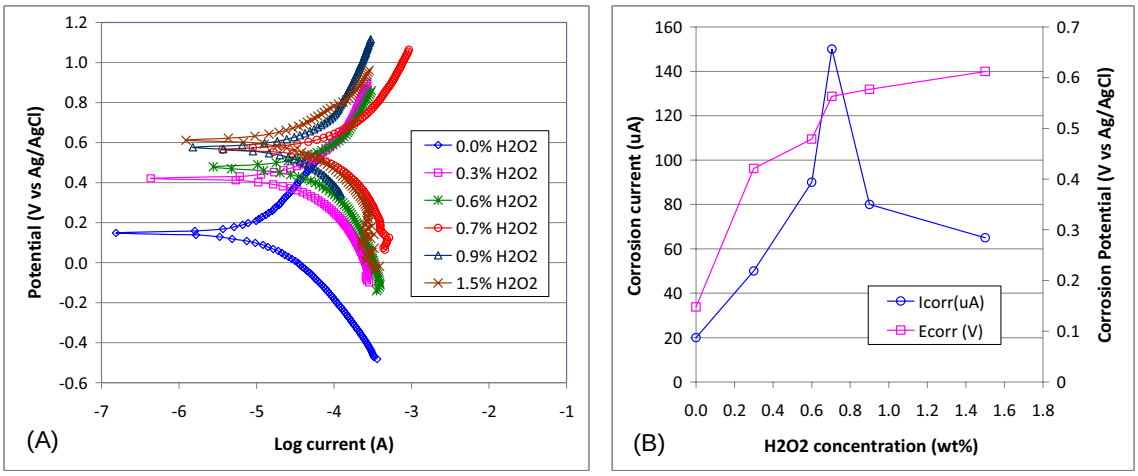


Figure 5. (A) Potentiodynamic polarization curves, and (B) corrosion potential and corrosion current for Cu in FCC1024M at 1:10 dilution containing different H_2O_2 concentrations.

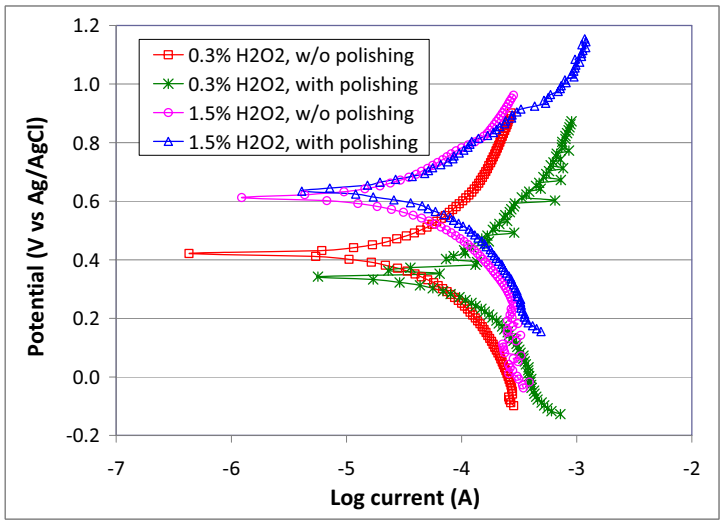


Figure 6. In-situ and ex-situ potentiodynamic polarization curves for Cu in FCC1024M at 1:10 dilution containing two different H_2O_2 concentrations.

ECS Transactions, 18 (1) 491-495 (2009)
10.1149/1.3096491 © The Electrochemical Society

Development of Copper CMP Slurry

Jianfen Jing, Chunxiao Yang, Xinyuan Cai, Danny Shiao, Shumin Wang

Anji Microelectronics (Shanghai) Co., Ltd., Shanghai, 201203, China

Abstract

Chemical mechanical planarization (CMP) is a critical technology for the copper damascene process. With the development of more advanced technology nodes, New challenges have emerged for the conventional CMP processes——the need for using lower down force while maintaining removal rate to avoid low-k/ultra-low k dielectric film damage and wide over polishing window. Also, surface defects especially corrosion related defects become more stringent as copper line becomes narrower and narrower. In this paper, we share our current efforts and associated results to develop the novel copper slurry for such CMP process. The slurries with different organic additives were evaluated for removal rate, dishing rate, electrochemical analysis and anti-corrosion capability. Based on these results, a novel copper slurry has been developed and it shows the excellent CMP performance.

Introduction

CMP is a key processing step in the copper Dual Damascene technique. As the progress of technology node, the industry moves to low-k dielectric which promises much faster devices with lower power consumption [1], copper CMP has become more stringent. The challenges include increasing requirement for copper dishing and erosion performance while maintaining high throughputs and low defects like corrosion, scratches and surface contamination.

To effectively planarize Cu surface, a balance between dissolution rates of the surface using an oxidizing agent, formation of thin passive layer with help of an inhibitor and removal of excess copper with help of complexing agents is needed [2]. In this study, we focus on the effect of other additives on the removal rate, topography and corrosion control capability.

Experimental

1. Polishing experiments

The polishing experiments were carried out using a bench top polisher, with IC1010 pads (Rodel Inc.) and 3M A165 diamond disk. Cu, Ta and Teos blanket wafers and Sematech 854 patterned wafer were used for all the polishing experiments. The down force pressure was 2psi/1psi for polishing head and 6lbs for in-situ conditioning disk. The platen and head rotation speed were set at 70rpm and 60rpm, respectively.

2. Electrochemical studies

Potentiodynamic polarization tests were carried out by using Electrochemical Instrument CHI660b. The Cu electrode (working electrode) used in these experiments

Ex. 1016, Page 519

was attached to a Teflon rod. Saturated calomel electrode and a platinum electrode were used as the reference and counter electrodes, respectively. The copper electrode was first allowed about 900s to attain a stable open-circuit potential. Tafel plot was then obtained by scanning the potential at a rate of 10mV/s from −1.0V to 1.0V.

Results and Discussions

1. Blanket removal rate and selectivity

The base slurry was formulated with colloidal silica, a complexing agent, a conventional inhibitor and hydrogen peroxide to provide high Cu rate. Several additives were added to the slurry to be evaluated. Fig.1 shows the Cu blanket removal rate of slurries with different additives. All of the additives (A~E) will decrease Cu RR, but they have different effects at different down force. Additive B, C and D show a constant decrease in RR for polishing down force. This indicates that they will lower dishing but lose throughput at the same time. Additive E is a good inhibitor but it drastically decreases RR. Additive A can suppress Cu RR at low down force more than at high down force. This indicates that additive A can inhibit chemical reaction but the passivation film can be removed easily by the high mechanical force. These characteristic of additive A is beneficial to lower dishing while maintaining throughput.

Fig. 2 shows the additive A concentration effect on Cu and Ta RR. Cu RR at high down force decrease smoothly with additive A concentration, while Cu RR at low down force will significantly decrease at high additive A concentration. This may result in Cu residue issue after P2 polishing. The slurry has low Ta RR and high selectivity.

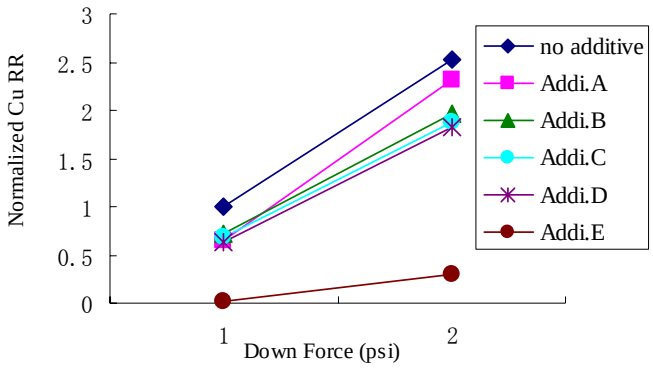


Fig. 1 Removal rate for slurry with different additives

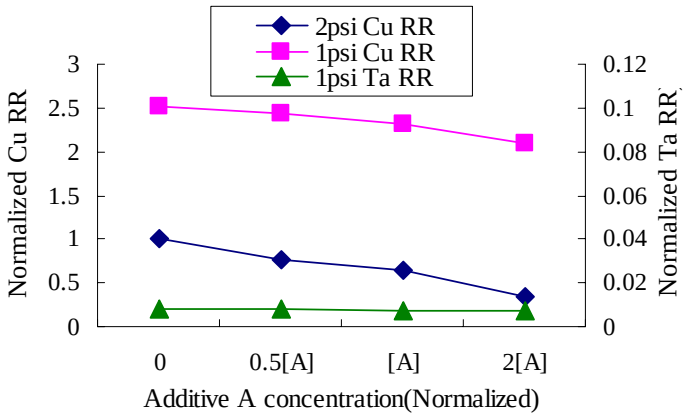


Fig. 2 Removal rate as a function of additive A

2. Dishing performance

Cu line dishing performance at over-polishing was investigated using slurries with and without additive A. The dishing depth was measured by scanning AFM at 20s, 25s, 30s and 40s over polish. Fig. 3 shows the dishing result. Compared with the slurry without additive A, the slurry with additive A results in lower initial dishing and dishing rate with the over polishing (OP) time.

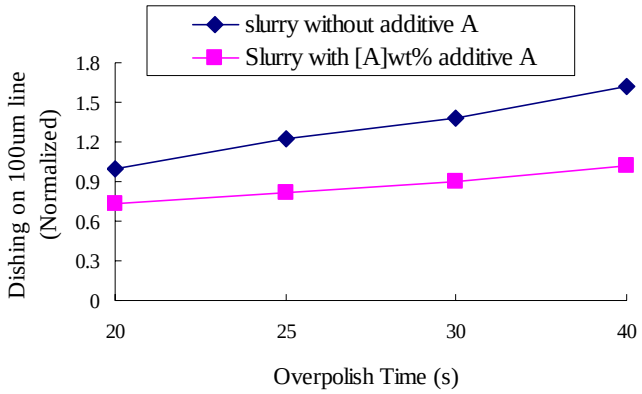


Fig. 3 The dishing performance at over polishing time for slurry with and without additive A

3. Corrosion control capability

Corrosion is a killer defect in CMP process. A good copper CMP slurry should have strong corrosion control capability. Electrochemical technique has been a widely used method for corrosion evaluation in semiconductor manufacture [3]. The protection film corrosion resistance can be extrapolated from electrochemical data. Patterned wafer soaking test (see the SEM images from this soaking test) also be used to investigate the corrosion control capability of slurry in this work.

Electrochemical analysis

Fig.4 shows the Tafel plots for copper in slurry with and without additive A. The corrosion potential (Ecorr) and the corrosion current density values (Icorr) were calculated by linear extrapolation of the cathodic and anodic branches of the Tafel plots and showed in Fig. 5. The corrosion potential (Ecorr) increase and corrosion current (Icorr) decrease after additive A addition indicating that a protection film formed. However, excessive addition of additive A didn't show more protection capability. Bias-potential versus anodic current also shows the same trend (Fig. 6).

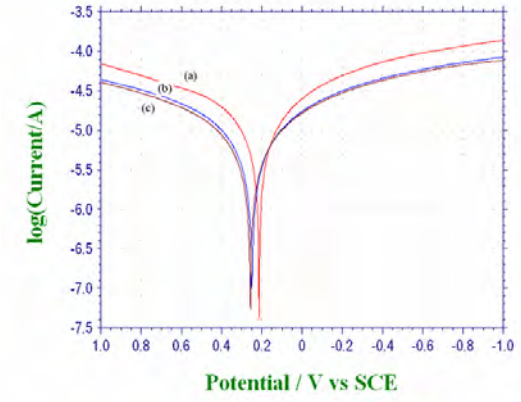


Fig. 4 Tafel plot for slurry with (a) 0 additive A; (b) [A] wt% additive A; (c) 2[A] wt% additive A

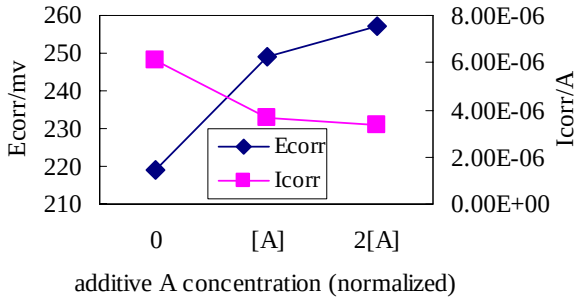


Fig. 5 Ecorr and Icorr as a function of additive A

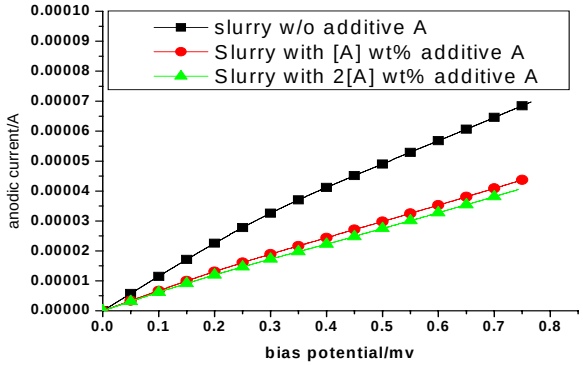
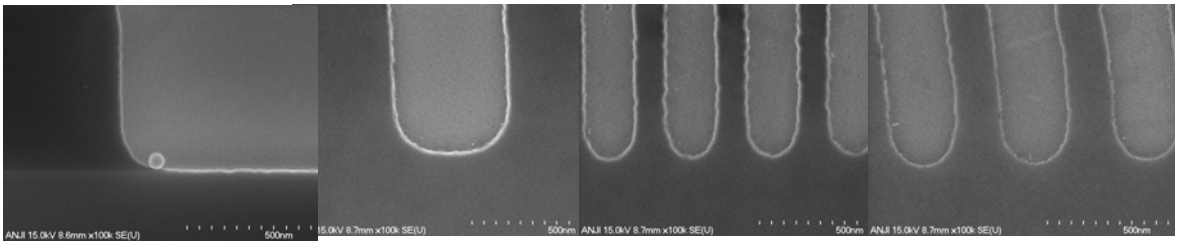


Fig. 6 Bias potential versus anodic current for slurry with different additive A concentration (normalized)

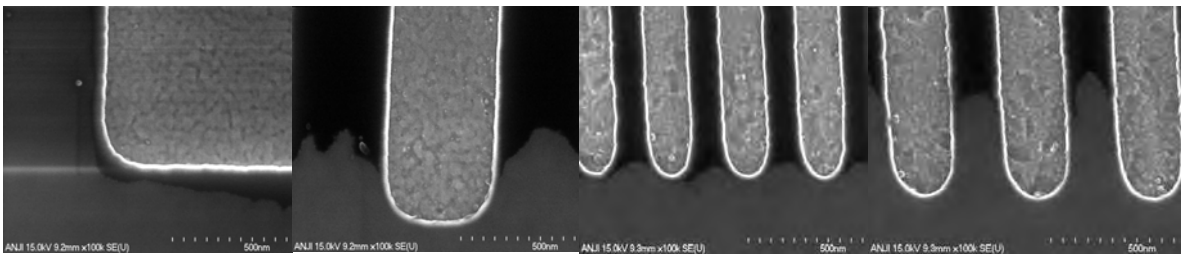
Patterned wafer soaking test

Patterned wafers are polished by slurry with and without additive A, and then soaked into the slurry for 30mins respectively. Then the surface was inspected by SEM.

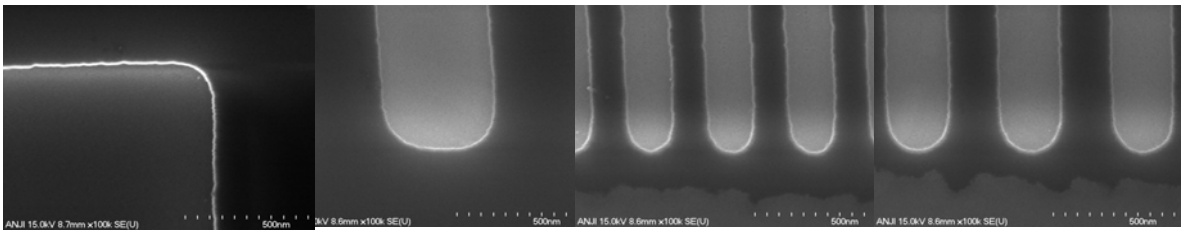
As shown in Fig. 7, fine surface on the inspected area is obtained by CMP process using slurry with and without additive A. However, corrosion is observed on the metal line after soaking in the slurry without additive A. The result illustrates that slurry with additive A has stronger corrosion control capability than that without additive A.



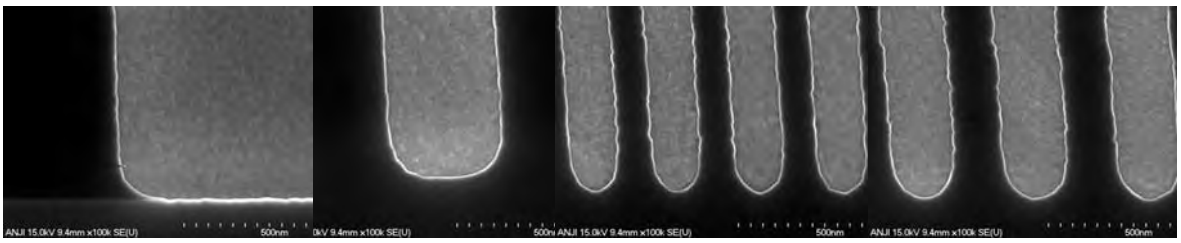
(a) post polishing with slurry w/o additive A



(b) post soaking with slurry w/o additive A



(c) post polishing with slurry with additive A



(b) post soaking with slurry with additive A

Fig. 7 patterned wafer SEM picture after polishing and soaking

Conclusions

In this work, several additives are evaluated in the base copper CMP slurry for decreasing dishing and dishing rate while maintaining high removal rate. Additive A was found to have suitable characteristics. It can improve slurry's capability to inhibit corrosion, lower dishing and dishing rate while maintaining high Cu RR at high down force. The slurry can meet the requirements of the industry.

Acknowledgments

The authors would like to thank all members of Anji microelectronics Co., Ltd. for their kind support.

References

1. M.R. Oliver, Chemical Mechanical Planarization of Semiconductor Materials, Springer, New York, 2004, p. 1.
2. P.B. Zantye, A. Kumar, A.K. Sikder, Mater. Sci. Eng. R R45 (2004) 89.
3. V. Brusic, M. A. Frisch, B. N. Eldridge, F. P. Novak, F. B. Kaufman, B. M. Rush, and G. S. Frankel, J. Electrochem. Soc., **138**, 2253 (1991)

ECS Transactions, 18 (1) 497-501 (2009)
10.1149/1.3096492 © The Electrochemical Society

A New Kind of Technology for Post CMP Cleaning in IC Fabrication

Li Wei-wei[a]

[a] Department of Information Engineering, Heibei University of Technology, Tianjin, 300401, China

Abstract

In this paper the forming course and mechanisms of contaminations were systematically analyzed. A new mechanism model has been established about contaminations on wafer surface. According to this model a new kind of cleaning method was putted forward. Block copolymer is used as cleaning solution coordinated by infiltration and corrosive under megasonic. The new cleaning method has three advantages: First, the adsorptive state of particles on wafers is effectively controlled. The efficiency and effect of megasonic particle removal is greatly improved; Second, it has the functions of chelating and non-volatile organic base. It can remove many kinds of metal ions; Third, it can remove organic substance because of the same structure. This new cleaning method can remove particles, metal ions and organic substances together. Its cleaning effect is good without environment pollution. So the new leaning method can be used in commercial production safely.

Introduction

With the development of semiconductor technique, there are more than 10 hundred million devices integrated on wafer surface. To improve circuit speed and reduce power consumption multi-layer wiring must be achieve. Chemical mechanical planarization(CMP) is the only best method achieving global planarization of a variety of material with different characteristic[1-6]. But the process of CMP may bring about many contaminations on wafer surface, including slurry particles, any particles from the polished material, and chemical contaminations from the slurry[7]. The pollutants can be classified to three kinds as particles, organics and meal ions. These pollutants must be removed after CMP, otherwise it will destroy following procedure or result in device failure[8]. At present, the requirement of cleanliness on wafer surface has been improved greatly. We are facing removing contamination with nano-size. Traditional cleaning method removing particles on wafer surface is using brush scrub[9-10], which is efficient but easy to scratch. Meanwhile, it is inefficient to the removal of organics and metal ions, and the scrubber is expensive and vulnerability. Post-CMP cleaning is very important to IC fabrication so we must remove all of the pollutants in one cleaning process after CMP[11-13].

The main reason of particle adsorption is that the vertical chemical bond was destroyed in the process of CMP, so the impending bonds came into being. These impending bonds adsorb all kinds of impurities easily to lower self-energy[14-18]. The impurities adsorb on the silicon in form of chemistry which is very difficult to remove. The smaller the particle has the easier the adsorption, because of contact ratio. We should use a kind of polyether

Ex. 1016, Page 525

with large molecule to adsorb on wafer surface prioritized. This kind of substance will be removed from wafer surface because their molecules are very large. Meanwhile these large molecules can form protective coating preventing re-adsorption. Polyether can remove metal ions and organics by forming soluble chelate and dissolution.

Cleaning Principle and Resolved Project

Wet cleaning method is using solvents、acids and alkalis as the cleaning solution. After erosion, dissolution and chemical reaction, the silicon was rinse with cold and hot water so that the impurities can be wiped off. Polyether is used as cleaning solution coordinated by infiltration and corrosive under megasonic. This method can achieve good cleaning effect, especially it will not destroy environment.

Cleaning Principle

In the process of CMP, there are a lot of nano-particles emerge coming from slurry and wafer surface material. Superficial expanded because of the shrink of particle size, which makes the occupied atoms and groups increased badly. Meanwhile, there are many impended bonds on wafer surface because of mechanical friction. So these particles are easily adsorbed on wafer surface under the function of surface hydrogen bonds and other chemical bonds in order to lower energy. On the other hand, there are a large number of charges on particle surface because of the groups, so these particles are unstable, and tend to adsorbed each other. The decomposed particles from wafer surface have high activity, and the energy of surface atoms is much higher than the energy of internal atoms. So the particles are easily diffused on wafer surface and linkage with corresponding atoms, thus permanent adsorption engenders which is difficult to be desorbed.

The cleaning purpose is to remove the particles before the forming of chemical adsorption and protect the surface from secondary pollution. Polyether has surface activity, emulsifying properties and dispersion performance. Polyether molecules directed adsorbed on wafer surface, and hydrophilic group extent to aqueous phase and hydrophobic group adsorb on particle, therefore, solid-liquid surface tension is lower. Then the particles can be wetting completely. Polyether can embed and absorb on crack wall, then bring entropy repulsion and infiltration of hydration force, which weaken the adsorption force between particles and wafer surface, and contribute to particle removal. Polyether molecules adsorb on wafer surface firstly which prevent secondly particle adsorption.

Using Non-ionic Penetrant

Nonionic penetrant is using, the reasons are as follow: ①It would not ionize in the water and would not produce ion pollution; ②It can form easily cleaning adsorption and easy to wipe off because its molecules is big; ③ It can be adsorbed on the surface of alive solid. So the stability of substance would be improved greatly. The penetrant molecules can extent in the crack between particle and wafer surface, which can avoid the forming of linkage, then improve the effect of particle removal[2]; ④ The acid, alkali and dielectric can not effect it. According to all of the above, nonionic penetrant is used to cooperate with polyether, and improve each other's advantages. Meanwhile polyether as well as penetrant can remove organic pollution because its molecular structure is similar to organics.

Ex. 1016, Page 526

Using Chelating Agent

Chelating agent can remove metallic ions effectively. The reason is that a large number of metals especially transition mental can form chelate. Even Na^{+}, Li^{+} and Mg^{2+}, Ca^{2+} can form chelate with a certain kind of chelating agent. Chelate is a kind of complex compound with orbicular structure. Ions of complex compound can combine with electric ions with contrary sign. It can almost complete ionization in water as strong electrolyte. Amine chelate is used because it will not bring metal ions.

Experimental Process and Result Analysis

Solvent: polyether, nonionic penetrant, chelating agent and other auxiliary reagent and deionized water.

Washer: TCQ-250 cleaning machine

The PH of cleaning solution is adjusted to almost neutrality.

Organics removal

Organics on wafer surface will destroy device performance. Fig1 shows the appearance and spectrum analysis of wafer surface before cleaning. There are C and O element, which shows organic residual on wafer surface.

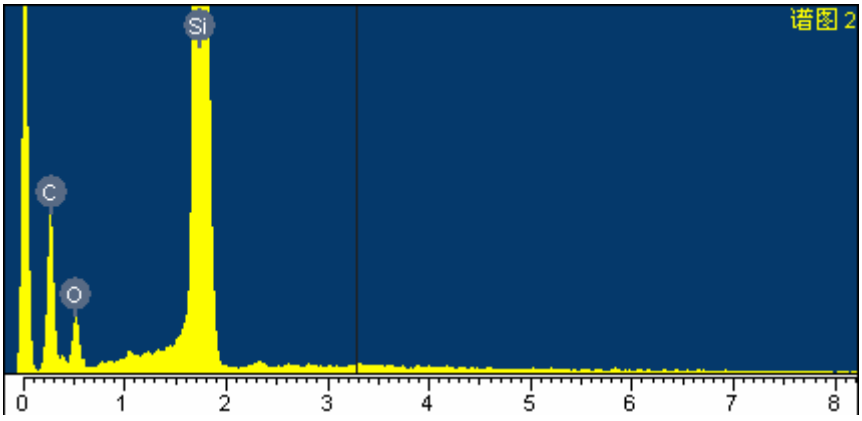


Figure 1. Spectrum analysis of wafer surface before cleaning

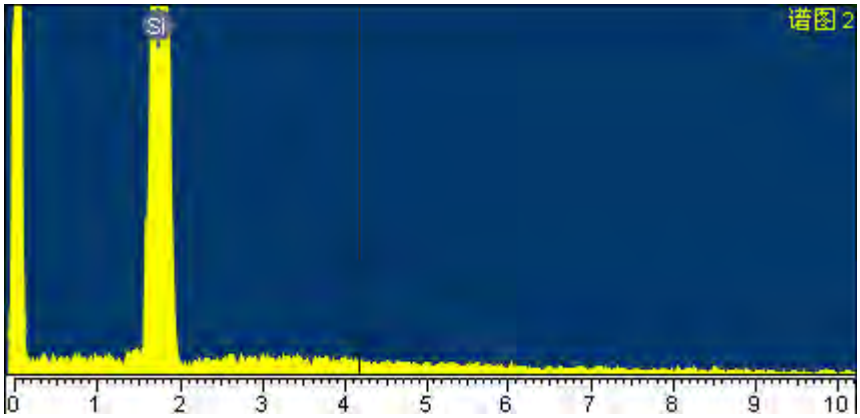


Figure 2. Spectrum analysis of wafer surface after cleaning

The organics have been removed from wafer surface after cleaning because of the function of polyether in cleaning solution. Polyether can remove organics because of its emulsification. The thermodynamic instability and increasing of interfacial area resulting from emulsification can be avoided because of the lowered surface tension. Meanwhile liquid drops are difficult to agglomerate due to the space barrier of adsorption layer.

Particles Removal

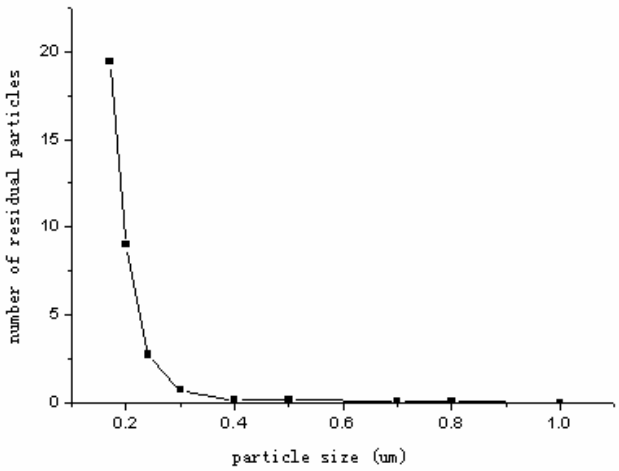


Figure.3 Number and diameter of residual particles after cleaning

Figure.3 shows number and diameter of residual particles after cleaning. It make us known that particles larger than 0.3μm is very little because of the function of large molecule polyether in cleaning solution. polyether molecules will combine with impendent bonds on wafer surface firstly as physisorption. Physisorption is intermolecular force so it is easy to break off. Protective coating will be formed after polyether molecules adsorbed on wafer surface, so it can also prevent secondary adsorption of particles. If particles has adsorbed on wafer surface, penetrant molecules can infiltrate to wafer surface and lift the adsorbed particles. Thus the adsorbed particles will desorp from wafer surface. To sum up, polyether plays an important role on particle removal.

Metal ions removal

The main metal taints are Cu, Fe and Cr on wafer surface. Traditional cleaning method can remove most of these metal ions. We use a kind of chelant with many chelate loops synthesized from three main components. This chelant can complex with more metal ions because it has more coordinate bonds than traditional EDTA. Table 1 shows metal ion residual using synthetic chelant.

Table 1 Metal ion residual using synthetic chelant

Metal ions	Synthetic chelant
Cu	$<2.0\times10^{-11}$
Fe	$<7.0\times10^{-10}$
Cr	$<1.0\times10^{-9}$

Conclusions

1. This cleaning method is innocuous and environmental. It can solve the key problem of post-CMP cleaning.

Ex. 1016, Page 528

2. The cleaning effect will be all right if we use compound solution of polyether, penetrant、chelant and DI water. It can remove organics, particles and metal ions all together in one cleaning without destroy wafer surface.
3. This cleaning method can cooperate with existing production facility and technics without modifying plant conditions. So it is easy to be accepted by production units.
4. The cleaning solution is nontoxic、non-pollution、easily formulated 、process simple. The effect is improved greatly because it can clean slices in great number. So it is suitable for industrialization.

References

1. T. hattori, V.J. Ruzyllo, R. Novak, Wafer cleaning, J. Electrochem. Soc. 97 (35) (1998) 1.
2. D. hymes, I. Malik, Brush scrubbing emerges as future wafer-cleaning technology, Solid State Technol. 7 (1997) 209-214.
3. Guthrie, W. L., Patrick, W. J., Levine, E, Jones, H. C., Mether, E. A., Houghton, T. F. Chiu, G. T. and Fury, M. A. IBM J. Res. Develop., 36, 845 (1992).
4. Ali. L., Roy, S. R, and Shinn, G., Solid State Technol., 37(10), 63 (1994).
5. Steigerwald S P. Murarka, Mechanisms of copper removal during chemical mechanical polishing [J]. J. Vac. Sci. Technol. 1995, B 13(6). Nov/Dec:2215-2218.
6. Ronald J, Gutmanm, Joscph M. Steigerwald, Chemical-mechamical polishing of copper with oxide and polymer interlevel dielectrics [J]. Thin Solid Films, 1995,270:596-600.
7. Liming Zhang, srini Raghavan, Milind Weling, J. Vac. Sci. Technol. B 17 (1999) 2248.
8. Liu Yuling, Liu Na, Cao Yang. Study on the state of adsorption on the silicon surface. Rare metal.1999,23(2):85~88
9. W.C.Krusell, J.M.de Larios,J.Zhang. Mechanical brush scrubbing for post-CMP clean. Solid State Technologe,1995,(7):109.
10. J.M.de Larios, M. Ravkin, D. Hetherington et al. There and back again. SEMICONDUCTOR INTERNATIONAL, 1996,(5):121~126.
11. Schaefer, D. M., Carpenter, M., Reifenberger, R., DeMejo, L. P., and Rimai, D. S., J. Adhesion Sci. Technol. V8, n3, pp197-210 (1994).
12. Schaefer, D. M., and Reifenberger, R. Mat. Res. Soc. Symp. Proc. V332, pp225, (1994).
13. Schaefer, D. M., M., Carpenter, Gady, B., Reifenberger, R., Demejo, L. P., and Rimai, D. S, J. Adhesion Sci. Technol. V9, n8, 1049-1062 (1995).
14. McFarlane, J. S., and Tabor, D., “Adhesion of Solid and the Effect of Surfacce Films,” Proc. Roy. Soc., A202:224 (1950).
15. Zimon, A. D., Adhesion of Dust and Powder, Plenum Press, New York, (1969).
16. Larsen, R. I., “The Adhesion and Removal of Particles Attached to air Filter Surface,” Am. Indust. Hyg. A. J., 19:265 (1958).
17. Bowden, F. P., and Throssel, W. R., “Adsorption of Wafer Vapor on Solid Surface,” Proc. Roy. Soc., A209:297 (1951).
18. Kordecki, M. C., and Orr, C., Archives of Environmental Health, I. (1960).

Ex. 1016, Page 529

ECS Transactions, 18 (1) 503-509 (2009)
10.1149/1.3096493 © The Electrochemical Society

Controlled DI Water Gasification System For Advanced Semiconductor Cleaning Processes

A. Xia, K. Niermeyer, G. Conner, R. Mollica

Entegris Inc., Billerica, Massachusetts 01821, USA

Driven by continually shrinking feature sizes and adoption of ever more fragile materials in IC manufacturing, it has become crucial to develop effective and low impact processes that are benign to features on wafers. Recently, there is increasing interest in using gasified DI water in photolithography, wet etch & clean, mask clean and CMP applications. In this paper, we describe the design and development of an automated in-line CO_2 gasification system using Entegris all PFA hollow fiber membrane-based contactors. The paper addresses key system design features to enable a responsive and seamless process with minimum system downtime. Capacity and control data demonstrating the versatility and robustness of the system will also be presented.

Introduction

To meet the cleaning requirements for wafers with features below 65-nm or incorporating emerging new materials, gasified DI water is playing an important role in advanced semiconductor cleaning processes. In particular, CO_2 is gaining much attention since dissolution of CO_2 in water is more than just a physical process. Unlike many gases, as CO_2 dissolves into water, it increases water's acidity by forming carbonic acid (H_2CO_3). Consequently, the dissociation of the acid produces more free moving ions in solution, which makes DI water more conductive. (Please see Eq. 1)

$$CO_2 + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons HCO_3^- + H^+ \rightleftharpoons CO_3^{2-} + 2H^+ \qquad \text{(Eq. 1)}$$

The conductive nature of the carbonated water makes it a better alternative than pure DI water to dissipate electrostatic charge during wafer rinsing. It is believed that a wafer free of electrostatic charge helps to reduce particle related defects. In BEOL cleaning or polishing processes, metal line corrosion occurs due to "pH shock" when an excess amount of hydroxyl ions is present. Using a low-pH CO_2 DI water solution eliminates the excess hydroxyl ions through a simple acid-base neutralization reaction.

One major challenge in DI water gasification is how to infuse DI water with small amounts of CO_2 in a controlled and consistent manner. The common practices to achieve low concentration of dissolved CO_2 include either diluting CO_2 with an inert gas before injecting the gas mixture into the membrane contactor or diluting highly gasified DI water with un-gasified water. However, both methods pose significant drawbacks. Mixing CO_2 with an inert gas introduces undesired gas species into the process. Diluting high concentration gasified water adds complexity in system design and control and proper mixing may not occur prior to dispense. Furthermore, both methods demand high consumption of either gases or water.

In this paper, we describe the design and development of an automated in-line CO_2 gasification system using Entegris all PFA hollow fiber membrane-based contactors. The system employs a novel method of direct injection of CO_2 into DI water without dilution to achieve and maintain ultra-low conductivity. The paper addresses key system design features to enable a responsive and seamless process with minimum system downtime. Capacity and control data demonstrating the versatility and robustness of the system will also be presented.

Main Applications of CO_2-DI Water

Due to its conductive and slightly acidic nature, CO_2-DI water rinsing has proved beneficial in various semiconductor cleaning processes including:

- Post-CMP cleaning
- BEOL cleaning
- Mask cleaning (1)
- Photoresist removal (1)

High Purity Water Gasification Technology

Gasification Using Membrane Contactor Technology

To make bubble-free gasified DI water, a perfluoroalkoxy (PFA) hollow fiber membrane contactor is used in a gasification system. The PFA membranes are potted into a PFA shell with PFA end caps. The all-PFA design delivers superior chemical capability, allowing the device to be used with a wide variety of fluids and gases for various applications. The hollow-fiber devices enable faster gas transfer rates than the conventional contactors, as the high membrane surface area-to-volume of such devices produces high mass transfer rates. Also, the hollow fiber module design is less prone to channeling that can compromise the performance of conventional equipment.(2) Figure 1 illustrates a typical counter-flow configuration in which CO_2 sweeps inside the hollow fiber (lumen side of the contactor) and DI water flows outside of the hollow fiber (shell side of the contactor). The hydrophobic membrane allows gas to freely diffuse into water, but prevents water from passing through the membrane into the gas side. The amount of CO_2 dissolved into water is controlled by adjusting the partial pressure of CO_2. The water electrical conductivity is directly proportional to the concentration of CO_2 in the water. Hence, in most applications, water conductivity is used as a measure of CO_2 concentration in water.

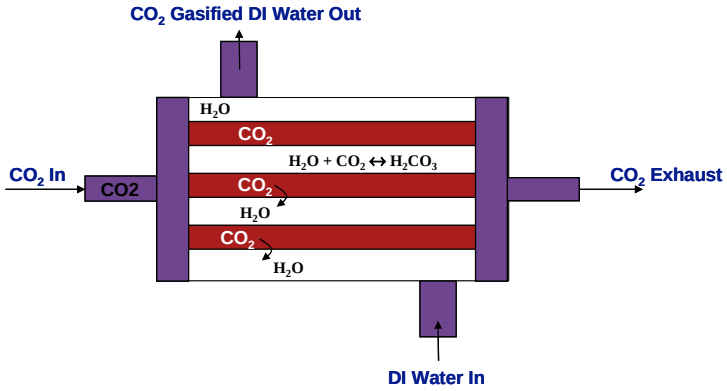


Figure 1. PFA gas-liquid contactor technology

Design of a CO_2-DI Water Gasification System

An ideal CO_2-DI water contacting system should at least comprise the following features to be beneficial to various semiconductor processes:

- Automatic conductivity control
- Optimized control loop with quick response and smooth control
- Direct CO_2 injection without using any gas or fluid mixing
- Wide range of conductivity
- Minimum gas/fluid waste and system maintenance for low cost of ownership
- Compact and efficient design for small footprint and reliability

CO_2-DI Water Contacting Efficiency

The main operating principle of a membrane contactor is governed by Henry's Law. Henry's law states that at a given temperature, the solubility of a gas in water at equilibrium is proportional to its partial pressure in the vapor-phase in contact with water. (Please see Eq. 2)

$$P = Hx \qquad \text{(Eq. 2)}$$

P = gas partial pressure; H = Henry's law coefficient; a function of temperature
x = concentration. of gas in water at equilibrium

Thus, in CO_2-DI water gasification process, to alter and maintain the amount of CO_2 dissolved in water, the system needs to adjust and control CO_2 pressure inside the membrane contactor. However, as certain rinsing applications require ultra-low conductivity (<10 μS/cm), it is not trivial to control low CO_2 pressure, forming dilute CO_2-DI water mixtures. As stated, the commonly used method involves diluting CO_2 with a neutral gas, such as N2. The neutral gas acts not only as a dilutant, but also as a carrier gas to quickly disperse small amounts of CO_2 into DI water. Depending on how low the conductivity is, a significantly large amount of diluting gas may be required, as shown in Table 1. The disadvantages of using such method are high total gas consumption and addition of undesirable gas species in the process. In addition, the method introduces a greater chance of outgassing to occur and bubble formation. We have developed a novel method of making extremely dilute CO_2-DI water mixture by direct injection. This new approach does not require any type of gas or fluid mixing. Combined with the high contacting efficiency of the device, this direct injection method eliminates the need for a diluting gas and lowers total gas consumption. Figure 2 displays typical CO_2 consumption vs. DI water flow rates at room temperature for conductivity set points of 6 μS/cm, 20 μS/cm and 40 μS/cm, using an Entegris all-PFA membrane contactor. In addition, the new method is able to quickly and uniformly distribute small amounts of CO_2 inside the contactor, which results in fast response time.

TABLE 1. Gas Consumption Comparison between CO_2 / N_2 Mixing Method and Entegris' Direct Injection Method

Method	CO_2 Consumption	N_2 Consumption	Target Conductivity	DI Water Flow Rate
Direction Injection	0.001 SLM	0	1 μS/cm	1 LPM
CO_2 Dilution with N_2	0.02 SLM *	32 SLM *	1 μS/cm	1 LPM

* A CO_2/N_2 flow ratio of 1 : 1600 needs to be maintained in order to achieve 1 μS/cm of conductivity

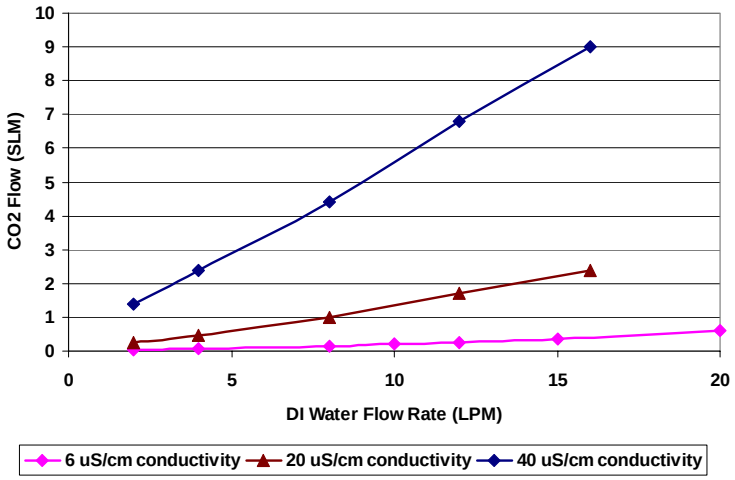


Figure 2. CO_2 Consumption vs. DI water flow to maintain 6µS/cm, 20µS/cm and 40µS/cm conductivity set point at 25°C, using a single contactor

Since different processes may require different CO_2 concentrations in water, it is important that the gasification system be able to deliver a wide range of conductivity for various water flow rates. Table 2 shows the typical minimum and maximum conductivity that a single membrane contactor can achieve at 1 LPM and 20 LPM water flow rates. By utilizing the unique method developed, a small amount of CO_2 can be directly injected into the water to maintain a conductivity as low as 0.5µS/cm, without using any mixing. For applications demanding high CO_2 concentrations, the system is able to produce water conductivity as high as 80 µs/cm for a water flow of 1 LPM, and 40 µS/cm for a water flow of 20 LPM. The maximum achievable conductivity decreases at a given CO_2 pressure as water flow rate increases due to the contacting efficiency becoming residence time limited. Higher conductivity can be achieved in high DI water flow applications by the use of multiple Entegris pHasor II contactors, effectively increasing the residence time.

TABLE 2. Typical conductivity range for 1 LPM and 20 LPM DI water flow rate in a single Entegris pHasor II membrane contactor unit at 25°C and under CO_2 pressure up to 40 psi

DI Water Flow Rate	Minimum Conductivity	Maximum Conductivity
1 LPM	0.5 µS/cm	80 µS/cm
20 LPM	0.5 µS/cm	40 µS/cm

System Conductivity Control Performance

As the industry moves towards single wafer processing and multiple-chamber cluster tool configuration, dispense cycles are shortened to maintain throughput, and process recipes become more complicated to accommodate increasing tool design complexity and functions. As a result, advanced cleaning steps demand a broad range of water flow and fast flow rate changes. Furthermore, concentrations of carbonated water (conductivity) still need to be tightly controlled and maintained to ensure a non-disruptive and stable process. The process complexity combined with stringent process control imposes a series of challenges on system conductivity control. Hence, control loop design is a critical step in designing a production-ready gasification system. An optimized control loop should not only stabilize the process during gradual changes, but also minimize

deviation and provide quick recovery during drastic flow rate swings. We have developed a PID-based conductivity control loop that is capable of handling various flow rate change schemes:

Gradual Water Flow Rate Changes

As shown in Figure 3, the conductivity is well within +/- 5% of the target conductivity of 6 µS/cm as the water flow rate changes 1LPM every 30 seconds between 8 – 12 LPM.

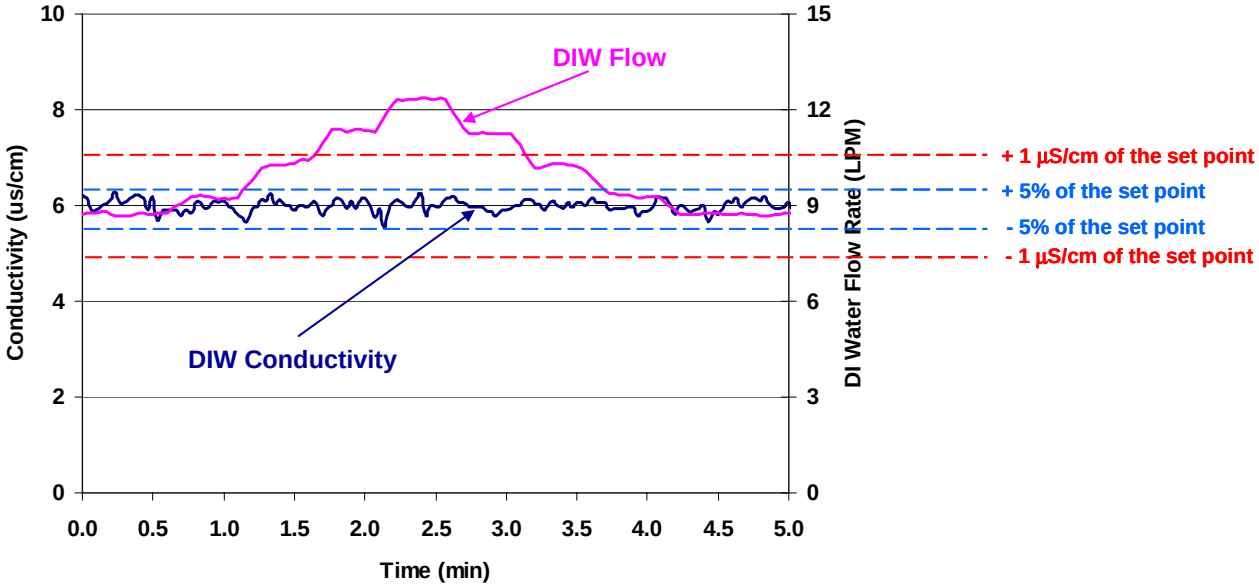


Figure 3 Conductivity vs. Time as water flow rate changes 1 LPM every 30 seconds between 8-12 LPM (Conductivity set point = 6 µS/cm; water temp. = 24°C)

Figure 4 simulates two back-to-back wafer runs, with 15-second wafer transfer time between each run. Each run consists of a 2 LPM change in water flow rate every 30 seconds between 2 LPM and 16 LPM. During a 15-sec wafer transfer, water flow rate stops and CO_2 flow shuts off. During each run, the control loop is able to maintain the conductivity within 5% of the set point. As the next run starts, the conductivity level recovers to the set point within seconds. Throughout the two runs including the idling during wafer transfer, the conductivity level never exceeds +/-10% of the set point.

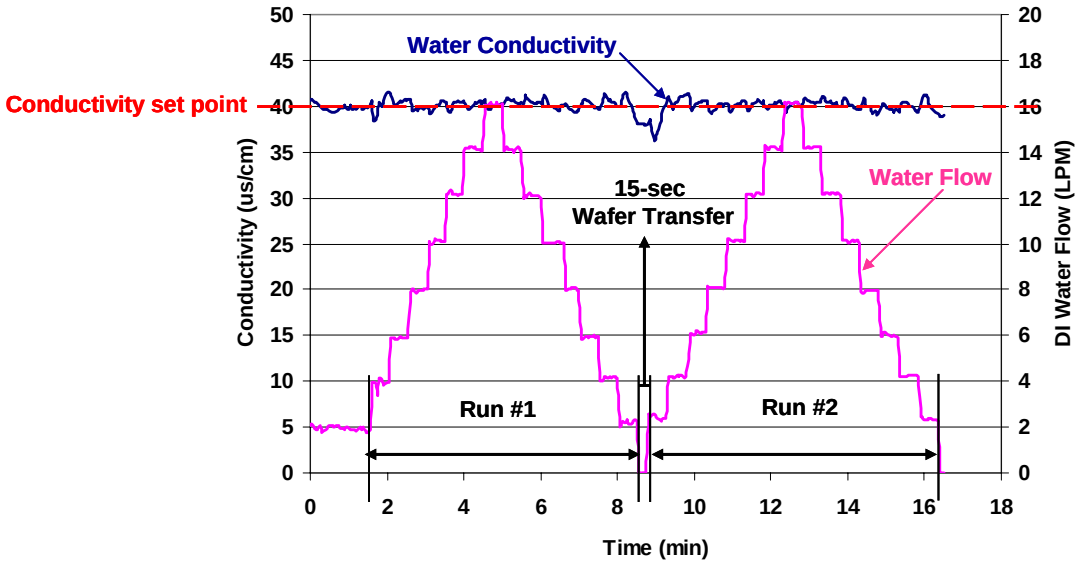


Figure 4 Conductivity vs. Time as water flow rate changes 2 LPM every 30 seconds between 2 – 16 LPM (Conductivity set point = 40 µS/cm; water temp. = 24°C)

Drastic Water Flow Rate Changes

Drastic water flow rate changes are not uncommon in multi-chamber processes. Depending on the magnitude of water flow rate changes, sometimes a traditional PID control algorithm might not be sufficient to deliver the acceptable response and stability. During system development, we have realized that additional control optimization is needed to minimize the conductivity overshoot when water flow rate drops sharply. As water flow rate decreases, it takes longer for the downstream sensor to sense any changes in water conductivity. Simple PID controllers are not designed to account for transient delays. Therefore, a conductivity overshoot compensation feature is implemented to minimize conductivity deviation during larger water flow decreases. Such a compensation feature is not necessary for undershoot offset since undershoot occurs when water flow rate increases, in which case sensing lag is not an issue. Figure 5a and 5b compare the amount of overshoot with and without compensation. When no overshoot compensation is used, a 20% deviation from the conductivity set point is observed as water flow decreases from 16 LPM to 2 LPM (Figure 5a). When overshoot compensation is used (Figure 5b), only a 10% deviation from the set point is experienced for the same water flow rate drop.

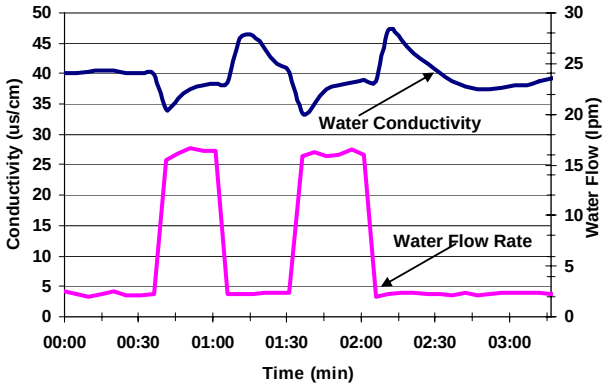


Figure 5a Water flow drops from 16 to 2LPM No Overshoot Compensation

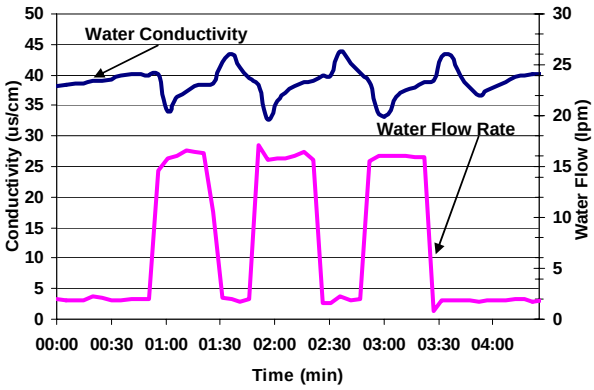


Figure 5b Water Flow Drops from 16 with to 12 LPM with Overshoot Compensation

Summary

An automated DI water gasification system using CO_2 direct injection has been designed. This system is capable of delivering a wide range of conductivity in varying DI water flow rates and can be easily scaled up to accommodate multiple membrane contactors for high water flow rate (>30 LPM) applications. Key design features include a novel method of direct injection of CO_2 into DI water for accurate flow conductivity response with a low cost of ownership and an optimized control loop that is highly responsive and effective in maintaining a stable conductivity for various water flow regimes. These design features are not just beneficial in delivering carbonated water, but are also applicable to other DI water-gas injection applications where accuracy and response is critical.

References

1. J. Papanu, *Mask Cleaning and Photoresist Removal Challenges*, Semiconductor International (April, 2007)
2. B. Parekh, *Emerging Applications of High Purity Water*, ULTRAPURE WATER (May/June 2008).

ECS Transactions, 18 (1) 511-516 (2009)
10.1149/1.3096494 © The Electrochemical Society

A Novel Slurry for DRAM Polysilicon CMP

Chen Wang[a], Yibin Cao[b], Chunxiao Yang[a], Haifeng Zhou[b], Jianfen Jing[a], Zhiping Xia[b], Jingji Tang[b], Jun Lin[b], Charles Chiu[b], Shumin Wang[a] and Chris Yu[a]

[a]Anji Microelectronics (Shanghai) Co., Ltd., Shanghai, 201203, China
[b] Semiconductor Manufacturing International Corporation, China

Abstract

High removal rate of polysilicon and high selectivity between silicon and oxide, are two important performance targets in DRAM chemical mechanical polishing (CMP). Traditionally, the polysilicon CMP is dominated by mechanical effects, owing to the relatively inert chemical property of silicon. We have shown that chemistry also played an important role in polysilicon polishing. In this paper, a novel polysilicon CMP slurry with an optimized chemistry is reported. It is specifically formulated for DRAM polysilicon polishing. It exhibits high removal rate, high selectivity (silicon/oxide), low defects and excellent surface quality.

Introduction

The chemical mechanical polishing (CMP) of polysilicon is an important technique to form polysilicon plugs, capacitor plate (such as for memory cell structures) or damascene lines in DRAM devices. The removal of polysilicon by CMP can reduce the recess in polysilicon plugs compared with conventional reactive ion etch (1). High removal rate of polysilicon and high selectivity between silicon and oxide, are two important performance targets in DRAM chemical mechanical polishing (CMP). Traditionally, polysilicon CMP is dominated by mechanical effects, owing to the relatively inert chemical property of silicon.

What sets Anji's slurry A developed in this work for polysilicon CMP apart from competitive slurries is its unique chemistry. Without chemistry, increasing the removal rate of polysilicon by mechanical effects (e.g. increasing abrasive particle loading or polishing down force) increases the removal rate of SiO_2 simultaneously, which leads to low selectivity (silicon /SiO_2). With proper chemistries added into the slurry, a remarkable increase in the polysilicon removal rate can be found, along with a low oxide removal rate, enabling outstanding selectivity.

In this study, we found a unique additive chemistry (called chemical B in this paper). It acts as not only an excellent accelerator, but also a stabilizer. The shelf life of the commercialized slurry A is more than 12 months, and it exhibits high removal rate even at high dilution ratio.

Experimental

Ex. 1016, Page 539

Polishing tests were carried out on 200 mm AMAT Mirra-Mesa using Rohm-Haas Politex high pad. Wafers were polished under down force 2 - 4.5psi. Platen speed and head speed are 70 rpm and 100 rpm, respectively. A slurry flow rate of 200 ml/min was used. Dielectric film thickness was measured on Thermalwave OP5240.

Results and Discussion

Chemical Effect in Anji Slurry A

Chemical B is a unique additive. It increases the polysilicon removal rate remarkably. As shown in Figure 1, with chemical B added into the colloidal silicon particle, about 3 times increase in removal rate can be found. This remarkable chemical effect can also be found in static etch rate (SER) experiment. As shown in table 1, the SER rate of polysilicon in un-diluted Anji slurry A is about 15 times higher than that of KOH solution.

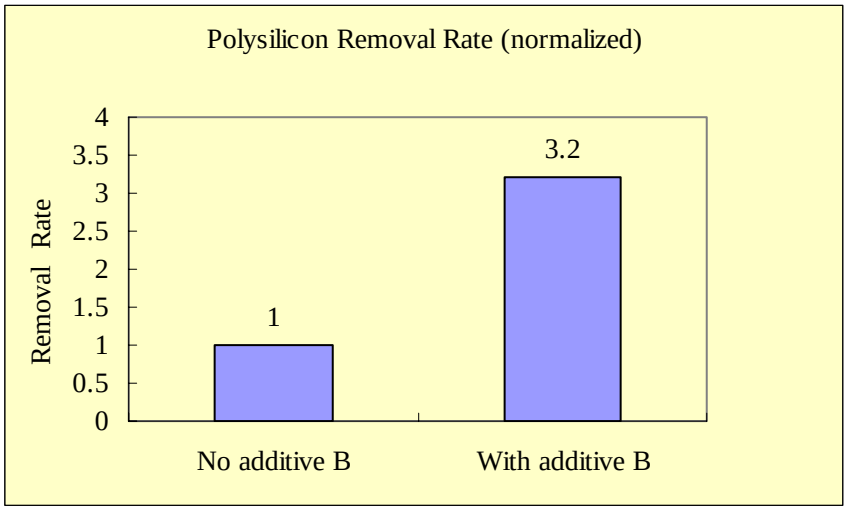


Figure 1. The effect of chemical B

Table1: Static etch rate of Anji slurry A

Slurry	Polysilicon static etch rate (A/hour)
KOH solution(pH 10.8)	246
Anji slurry A (un-diluted, pH 10.8)	3794
Anji slurry A (1:7 dilution, pH 10.8)	778

The mechanism of chemical B in anji slurry A is still under investigation. In conventional polysilicon CMP process, Lee and co-workers suggested that hydroxide bonds to the Si surface of the abrasives in the slurry and OH^- attaches to the polysilicon surface (2, 3). The silanol group induces high polarization and thus weakens the Si–Si back bonds. Then ambient H_2O or O_2 dissolved in the slurry directly break the weakened bonds to form either monohydrates or an internal oxide bridge, followed by dissolution by mechanical abrasion. Based on this mechanism,

chemical B might accelerate the reaction between OH^- and silicon surface. As a result, the polysilicon removal rate was increased remarkably.

Removal Rate vs Down Force

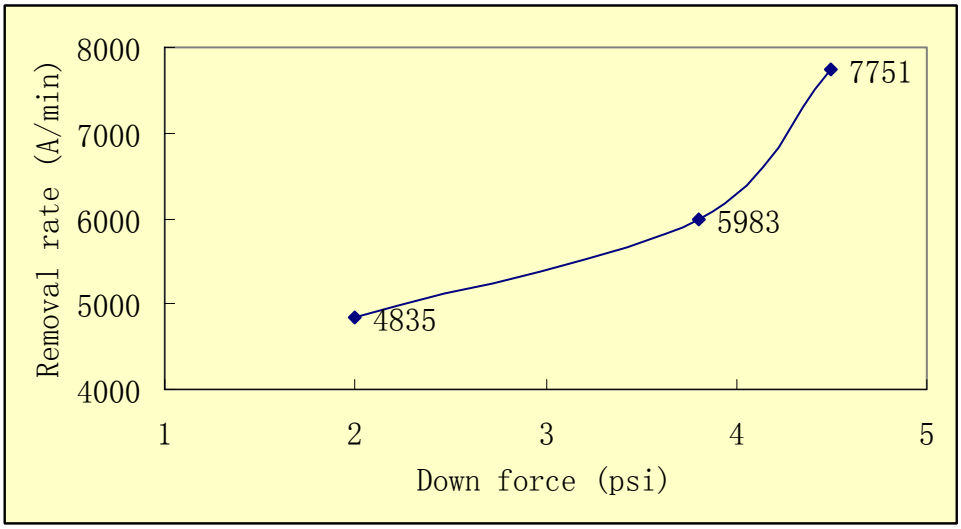


Figure 2. The removal rate as a function of down force on 8 inch wafer

High removal rate of polysilicon can be obtained by slurry A . The removal rate of slurry A increases with the pressure (down force). As shown in figure 2, polysilicon removal rate of slurry A can reach 7751 A/min at 4.5 psi.

Removal Rate vs Dilution Ratio

The dilution window of slurry A is wide. Table 2 shows the removal rate and selectivity under different dilution ratio. From 1:6 dilution to 1:8 dilution, the drop of removal rate is less than 7 percent. The highest selectivity of poly/Si_3N_4 (418) and polysilicon/oxide (208) was found at 1:7 dilution.

Table2: The removal rate and selectivity *vs* dilution ratio.

Dilution		**Removal rate(A/min)**			**Selectivity**	
ratio	**NU%**	**polysilicon**	**Si_3N_4**	**oxide**	**poly/Si_3N_4**	**poly/oxide**
1:6	2.48	6369	18	33.3	354	191
1:7	2.38	5983	14.3	28.7	418	208
1:8	1.91	5930	17.7	30	335	198

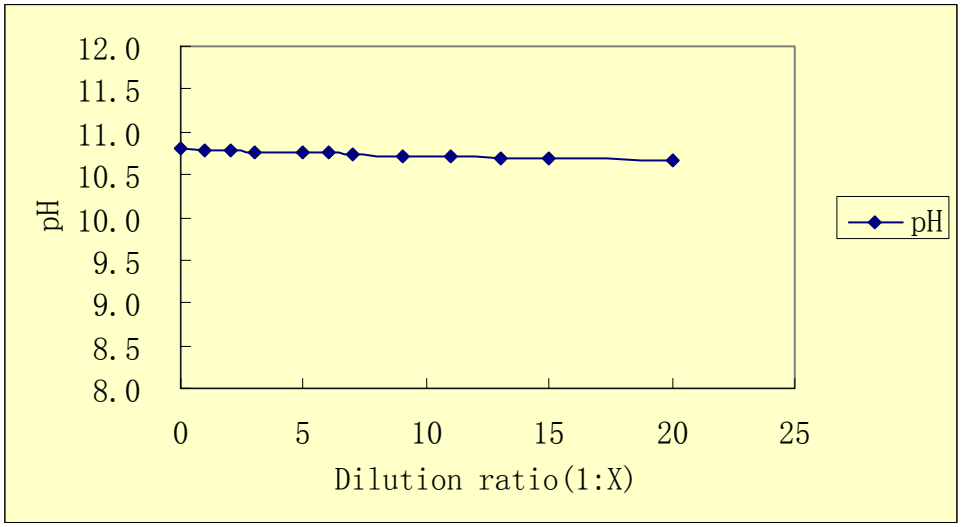


Figure 3. The pH as a function of dilution ratio

The excellent pH stability *vs* dilution (Figure 3) is expected to contribute to the stability of removal rate under different dilution ratio. The pH value drops less than 0.14 even after 1:20 DIW dilution.

Profile and Roughness

Figure 4 shows a removal rate profile of slurry A on a 8 inch polysilicon wafer. An average removal rate of 5983 A/min was obtained at 3.8 psi., and the wafer-edge profile is better than a competitive slurry which is widely used in DRAM polysilicon CMP.

Slurry A has an excellent surface finish, which is likely due to the proper balance between the chemical and mechanical components. The polysilicon wafer surface roughness is less than 0.6 angstrom under 3 psi (AFM), and might be further optimized by process .

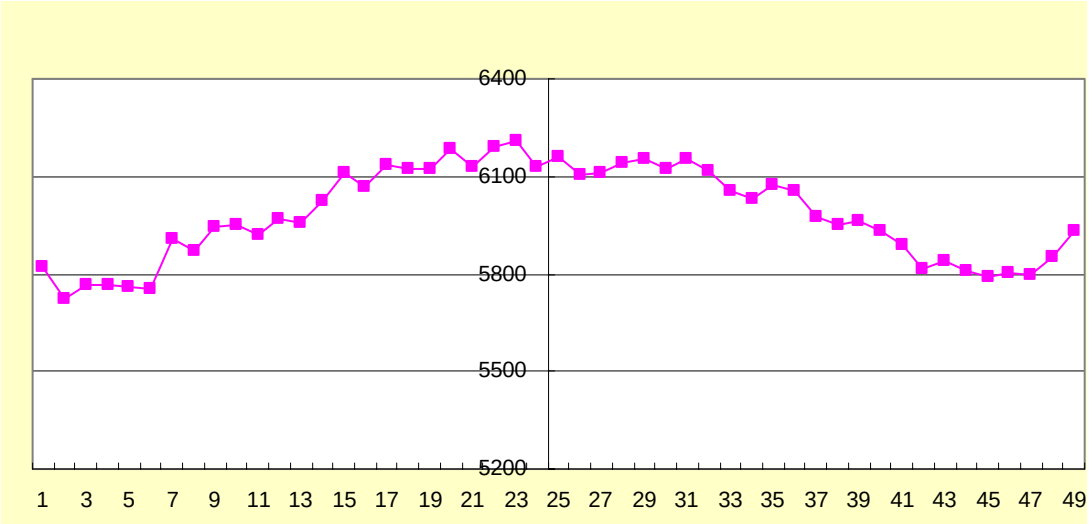


Figure 4. The removal rate profile of a 8 inch wafer (under 3.8 psi)

Mini-marathon on PE-Oxide

The performance of Slurry A was proved stable. The variation of removal rate, normal uniformity and particle defects were investigated by mini-marathon test on PE-Oxide wafers (Figure 5，6，7 ，under 3.8 psi, 1:7 dilution). The average removal rate of oxide is 28 A/min, and the average defects per 8 inch wafer is about 21. Evaluation results indicate that slurry A is suitable for polysilicon polishing, especially for DRAM polysilicon CMP.

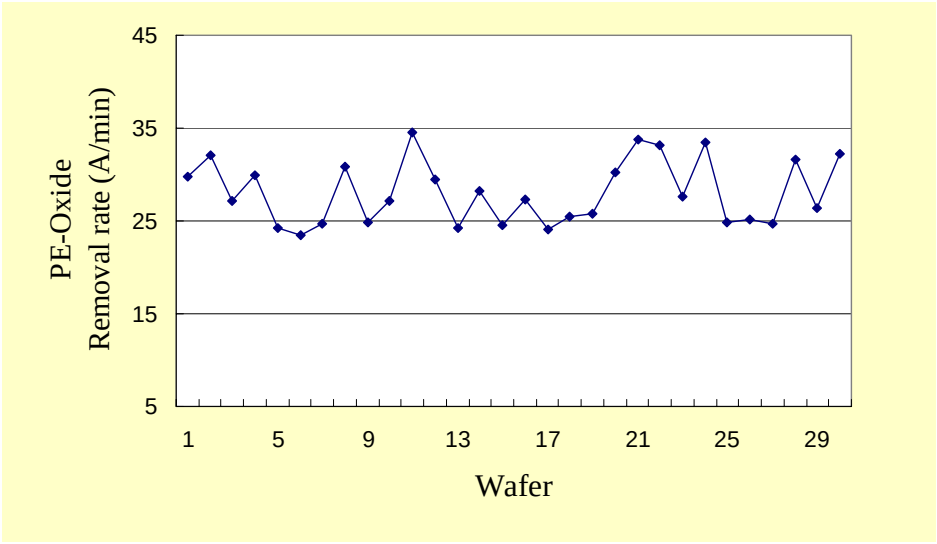


Figure 5. Mini-marathon on PE-Oxide (removal rate)

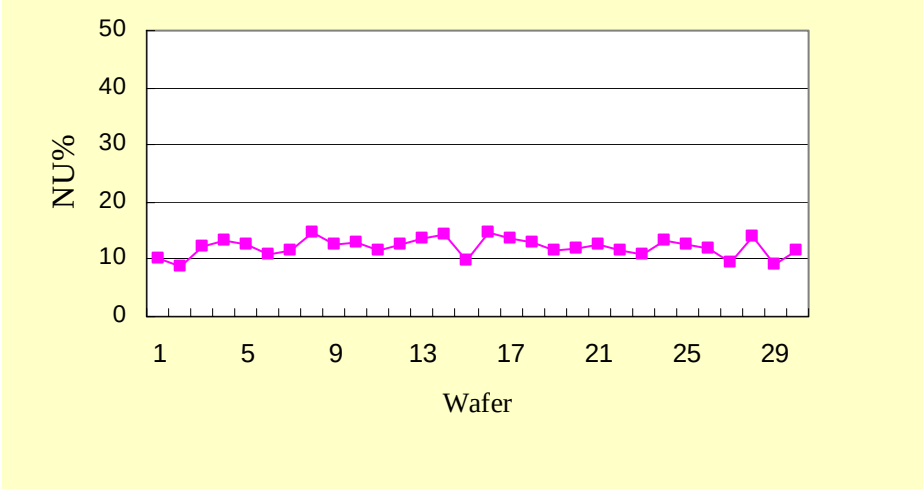


Figure 6. Mini-marathon on PE-Oxide (NU)

Ex. 1016, Page 543

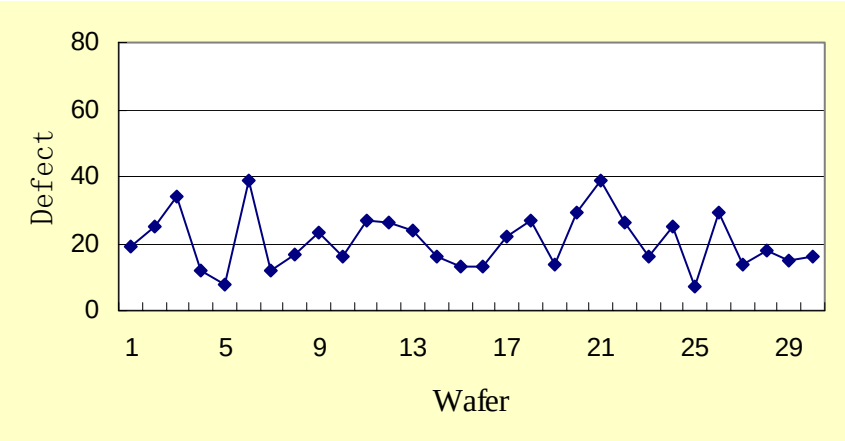


Figure 7. Mini-marathon on PE-Oxide (defect per 8 inch wafer)

Conclusion

A novel slurry for DRAM polysilicon CMP was developed and evaluated in this work. It has demonstrated that chemistry can also play an important role in polysilicon polishing. The removal rate of polysilicon and the selectivity to oxide was remarkably increased by the proper chemicals. 6000 A/min of polysilicon removal rate and 208 of poly to oxide selectivity can be obtained at 4.5 psi (1:7 dilution). Meanwhile, the slurry shows an excellent surface finish and low defectivity, making it ideally suited for DRAM polysilicon polishing.

Acknowledgments

The author would like to thank all CMP group members at Anji Microelectronics Co., Ltd, and Semiconductor Manufacturing International Corporation (SMIC) for their support.

References

1. J. Koh, D. Suh, D. Han, J. Kim, N. Park and S. Han, *VLSI and CAD, 1999. ICVC '99. 6th International Conference on* p.214 (1999).
2. G.j.Pietsch, G.S.Higashi and Y.J. Chabal, *Appl. Phys. Lett.*, **64**, 3115 (1994).
3. M.Y. Lee, H.G. Kang, M. Kanemoto, U. Paik and J.G. Park， *Jpn. J. Appl. Phys.*,**46**, 5089 (2007).

Ex. 1016, Page 544

ECS Transactions, 18 (1) 517-522 (2009)
10.1149/1.3096495 © The Electrochemical Society

A Matching Game For CMP Pads And Conditioners

F. Sun[a], J. Hawkins[a], J. Tsai[b], G. Chiu[b], and A. Naman[a]

[a] Cabot Microelectronics Corporation, Aurora, Illinois 60504, USA
[b] Cabot Microelectronics Corporation, Taiwan 302, R.O.C.

The relationship between a pad and a conditioner may be the most critical among all "consumables" in a CMP process. The effect of a pad conditioner on pad can be both short-term (instant) and long-term (stability). This paper compares the instantaneous effect of different conditioners on removal rate of oxide, tungsten, and copper CMP. The impact of conditioners in removal profile in copper polishing and rate stability (long-term) in oxide polishing are also presented. The pad wear rate by conditioners of different aggressiveness on Epic® D100 vs. conventional hard pad is compared and the benefits and risks of adopting a more aggressive conditioning on Epic® D100 pad vs. a less aggressive conditioning on conventional pad is discussed.

Introduction

There are two types of pads commonly used in CMP, hard pads (e.g., IC1000, Epic® D100) and soft pads (e.g., Politex, Fujibo). Hard pads are designed for greater wear resistance and thus longer pad life, but do not provide the defectivity performance of soft pads. Pad life should be defined as the shorter of two conditions, useful pad life - the time over which CMP performance is maintained, or natural pad life - the time required to cause significant wear to the pad such that grooves disappear. In order to achieve stable CMP performance over the life of the pad, hard pads require the use of diamond conditioners to maintain a constant surface "texture" and remove by-products. However, the use of diamond conditioners to stabilize texture accelerates pad wear and thus shortens the pad life. Until recently, the industry has focused on creating conditioning processes via the use of less aggressive diamonds or milder conditioning recipes in order to maintain performance stability while minimizing pad wear. Another approach that is showing promise is to engineer pads which are inherently more wear resistant through polymer selection and microstructure optimization. For example, the Epic® D100 (Cabot Microelectronics Corp.) shows 1~2x higher wear- resistance (or 1~2x lower wear rate) under same condition than most conventional hard pads used in CMP processes. Consequently, this pad will require a more aggressive conditioning in order to achieve a maximum useful pad life.

This paper presents the impact of conditioner aggressiveness on removal rates of oxide, tungsten, and copper polishing and the impact of conditioners on removal profile in copper polishing. The effect of conditioner aggressiveness on the pad wear rate of both the Epic® D100 and conventional hard pad is also presented. Finally, the benefits and risks of adopting a more aggressive conditioning on Epic® D100 pad vs. a less aggressive conditioning on conventional pad is discussed.

Ex. 1016, Page 545

Experimental

Polishing Experiments. All polishing experiments are conducted on Applied Materials' Mirra polisher (stand-alone). The slurries used were Semi-Sperse® SS25 (for oxide polishing), Semi-Sperse® W2000 (for tungsten polishing), and iCue®600Y75 (for copper polishing), all manufactured by Cabot Microelectronics Corporation. Epic® D100 pad was used for all experiments. Diamond conditioners were provided by Kinik Corporation (Taiwan) and 3M Company (USA).

Metrology and Pad Analysis. Oxide wafer thickness was measured on UV1050 (KLA-Tencor) using a 49-point diameter scan. Both tungsten and copper wafer thickness are measured on RS100, also using a 49-point diameter scan. Pad surface roughness was measured using Veeco® interferometer. Glazing value was obtained using Olympia optical microscope with the help of image-analysis software. Pad thickness was measured using Keyence microscope (cross-section) and mean wear rate was calculated based on post and pre thickness data.

Results and Discussions

Impact of Diamond Conditioner on Material Removal Rate

Diamond conditioning is used in CMP to counteract effects that are exerted on pad surface by many elements so that polishing performance can be stabilized throughout pad life. Elements that impacts pad surface (more specifically, surface "texture") during polishing include, a) temporary morphological change or deformation of surface "texture" as result of compression by the wafer and retaining ring under pressure, b) precipitated and/or embedded by-products and debris generated during polishing, and c) permanent loss of surface "texture", which is generally not recoverable through a normal conditioning process. **Figure 1** shows a typical material removal rate response to conditioning during oxide polishing with SS25.

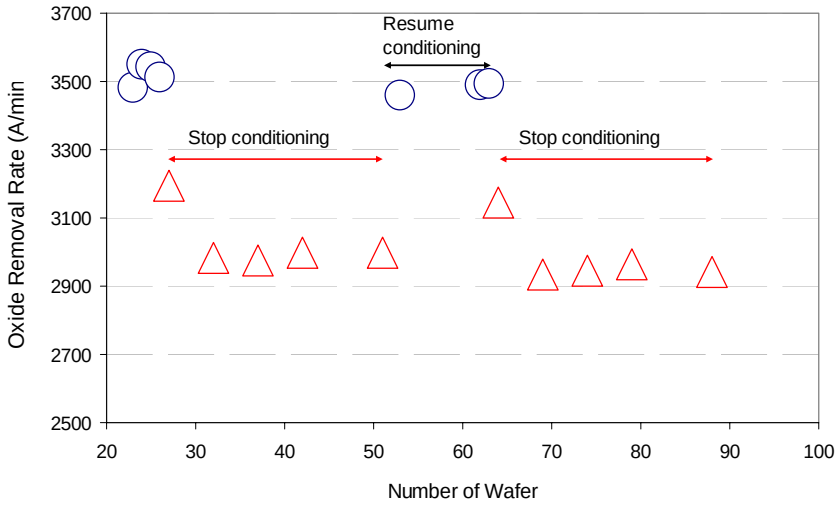


Figure 1. Response of material removal rate to conditioning during oxide polishing (**O** - with conditioning, **Δ**- w/o conditioning)

As can be seen, the rate drops quickly when conditioning stops. Since total polishing time is relatively short (25 wafers or ~ 0.5 hours) and removal rate is recovered immediately when conditioning resumes, rate drop in this case is probably a result of temporary morphological change of surface "texture" and/or embedded by-products and debris (a and b), rather than a permanent loss of some of surface "texture" (c).

Figure 2 compares the effect of two diamond conditioners of different aggressiveness on oxide removal rate over an extended pad life. Conditioner properties are summarized in **Table I**. A more aggressive conditioner (disk A) shows ~10% higher rate than its counterpart (disk B) throughout the entire pad life tested. The higher rate by the more aggressive conditioner is attributed presumably to a more favorable pad surface "texture" generated by disk A (instant or short-term effect).

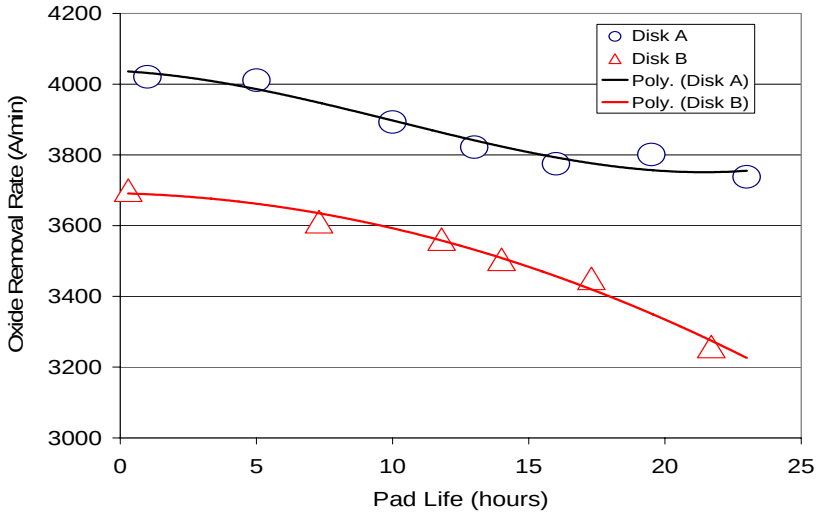


Figure 2. Effect of conditioner aggressiveness on oxide removal rate

In addition to the higher rate, the more aggressive conditioner (disk A) also shows a slower rate decay (~6%) over pad life compared to disk B (~12%). In fact, rate seems to "stabilize" at the end of test with the more aggressive conditioner, while the less aggressive conditioner (disk B) shows an even steeper drop.

Table I. Diamond conditioner property and corresponding pad surface characteristics

Disk	Diamond Type	Diamond Size (um)	Diamond Spacing (um)	Disk Shape	Disk Aggressive-ness rating [1)]	Mean Roughnes s (um)	Mean Glazing (%)	Mean Pad Wear Rate (um/hr)
A	3D	150	400	Donut	6	4.8	6	8.1
B	4D	180	220	Donut	3.5	6.1	9.4	6.4

1) normalized number based on pad wear rate provided by disk vendor

Table I also lists the overall pad surface roughness and glazing data. For surface roughness, a set of pads with identical rates, as shown in **Figure 2** but with much shorter pad life (~3 hours), were used to minimize possible influences from either insufficient pad break-in or excessive pad glazing, so that more accurate surface "texture" of the pad can be reflected. The pad surface roughness is relatively lower with the more aggressive conditioner (disk A). Surface glazing, an indication of both asperity flattening and debris

embedding, is also lower with the pad using the more aggressive conditioner (disk A) compared to its counterpart (**Table I** and **Figure 3**)

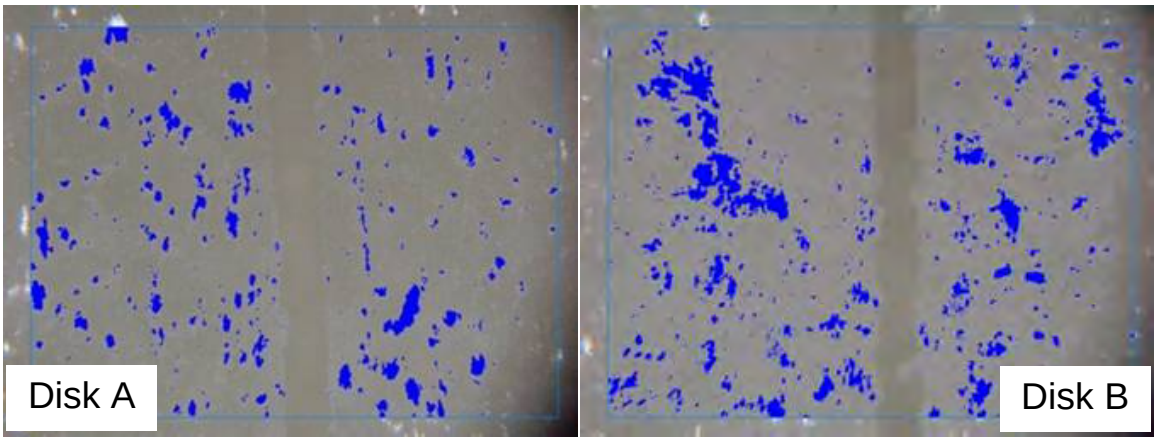


Figure 3. Pad surface glazing (after 21~23 hours of polishing)

Figure 4 compares the removal rate response to the conditioner aggressiveness for oxide, tungsten, and copper polishing. All conditioners were from the same vendor and disk aggressiveness rating is given based on pad wear rate provided by the vendor. Like the oxide case, tungsten removal rate also increased as the disk became more aggressive. Copper actually showed a slight decrease as the disk aggressiveness increased. Obviously, for oxide and tungsten, choosing more aggressive conditioners can benefit both in absolute rate (an instant effect), as well as improved rate stability (long-term effect) whereas in case of copper, a more aggressive conditioner may not boost copper rate while improving rate stability.

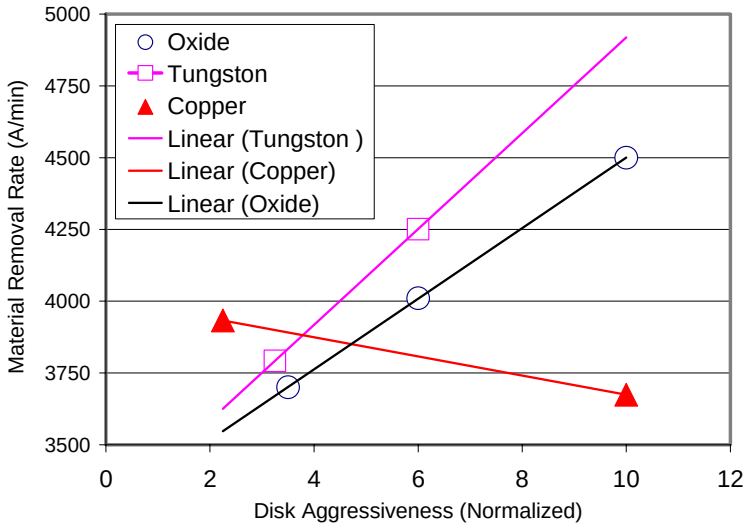


Figure 4. Comparison of removal rate responses to conditioner aggressiveness

Impact of Diamond Conditioner on Material Removal Profile

Process knobs such as the retaining ring pressure, the head sweep range, and even the slurry arm position are known to have an impact on removal profile across a wafer. Diamond conditioners can be another effective knob for obtaining a certain removal

profile. **Figure 5** shows an example of removal profiles in a copper polishing process by two different diamond conditioners from the same vendor (**Table II**). Under the exact same process condition, disk **C** produces a center-fast removal profile while the profile across wafer is almost flat for disk **D**. Such an impact of disk on polishing removal profile across wafers provides process engineers with some additional flexibility in adjusting removal profile.

Table II. Property of diamond conditioners

Disk	Diamond Type	Diamond Size (um)	Diamond Spacing (um)	Disk Shape	Aggressiveness
C	Semi-sharp	180	375	donut	more
D	Semi-sharp	250	500	full	less

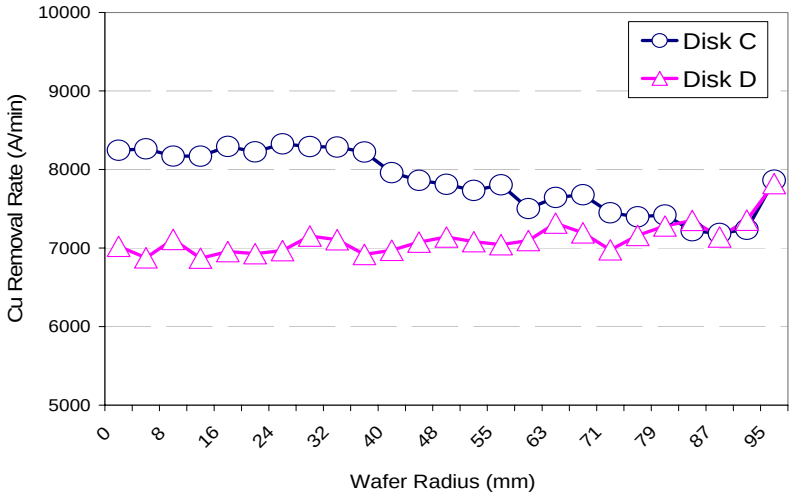


Figure 5. Effect of diamond conditioner on removal profile in Cu polishing

Balancing Pad Wear-Resistance to Conditioning Aggressiveness

Figure 6 compares the wear rate of pads using two groups of conditioners of varying aggressiveness. Epic® D100 shows a lower pad wear rate even with more aggressive conditioners (disk A and disk C) than conventional pad with less aggressive conditioners (disk B and disk D). This suggests one can use more aggressive conditioning and still possibly achieve a longer pad life on Epic® D100 because of its higher intrinsic wear-resistance.

Table III compares two different approaches for pad life extension. The biggest risk with the less aggressive conditioning approach is a potential for scratch "spike" and higher "sensitivity" to conditioning variation. It is well known that both scratches and pad stains tend to be worse at end of pad life, suggesting a deterioration of surface quality ("texture' and cleanness) over time. As technology nodes advance, this may pose an even bigger challenge because of lower tolerance for scratches and variations. The option with lower wear pad and more aggressive conditioning could provide an edge in terms of performance stability, robustness to conditioning variation, as well as real (useful) pad life, which is defined by how long the pad actually lasts in use, rather than a premature removal because of a spike in defects or a shift in removal rate or profile.

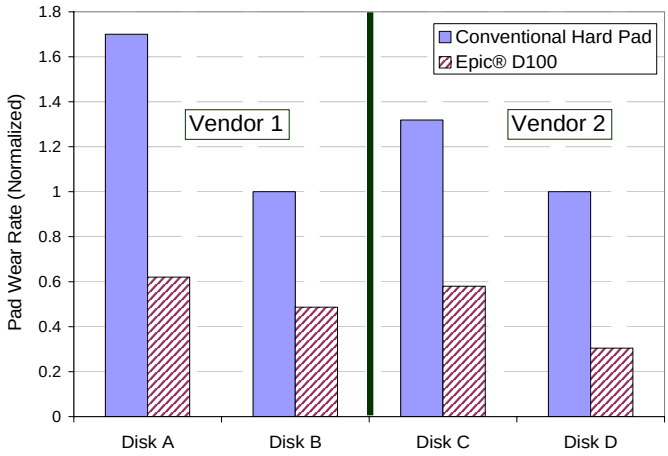


Figure 6. Comparison of relative pad wear rate by different conditioner

Table III. Comparison of two techniques for pad life extension

	Approach A	Approach B
Technique	Lower Wear-resistant Pad + Less Aggressive Conditioning	Intrinsically More Wear-resistant Pad + More Aggressive Conditioning
Pros	1. longer disk life (potential)	1. more consistent and clean surface 2. less "sensitive" to conditioning variation 3. extended real (or useful) pad life
Cons	1. potential rate decay and profile shift 2. higher risk of by-products/defects 3. more "sensitive" to conditioning variation	1. shorter disk life (potential)

Conclusions

Diamond conditioners have both a short-term (instant) impact and a long-term effect on pad performance. The sort-term impact is reflected by (absolute) material removal rate while the long-term effects are shown by monitoring performance stability over pad life. To achieve a stable performance, the pad surface texture needs to be maintained consistent and remain as clean as possible. This can be accomplished through a more aggressive conditioning. An aggressive conditioning will also benefit defect performance, especially when pad life is near or at an end because the surface has probably changed and by-product debris has accumulated to its worst level (if conditioning is insufficient). However, aggressive conditioning is not always accepted to end users because of enhanced pad wear and shorter pad life. This perception does not apply to a pad designed with intrinsically more wear-resistant polymer composition and structure like the Epic® D100. More aggressive conditioning can be (and as matter of fact should be) used on Epic® D100 to achieve a stable performance over an extended and “useful” pad life.

Acknowledgements

The authors would like to thank K. Brooks, M. McAleer, S. Liljegren, and R. Vandyke for providing technical assistance in polishing, metrology, and pad analysis, and D. Khanna, E. Remsen, J. Parker, and P. Feeney for meaningful discussions. Kinik Corporation (Taiwan) and 3M Company (USA) are acknowledged for providing all conditioners used in this study. Thanks also to Cabot Microelectronics Corporation for permission to publish this work.

ECS Transactions, 18 (1) 523-528 (2009)
10.1149/1.3096496 © The Electrochemical Society

Polycrystalline Diamond (PCD) Shaving Dresser: The Ultimate Diamond Disk (UDD) for CMP Pad Conditioning

James C. Sung[a,b,c], Cheng-Shiang Chou[a], Ying-Tung Chen[d]
Chih-Chung Chou[a], Yang-Liang Pai[a], Shao-Chung Hu[a], and Michael Sung[e]

[a] KINIK Company, 64, Chung-San Rd., Ying-Kuo, Taipei Hsien 239, Taiwan, R.O.C.
[b] National Taiwan University, Taipei 106, Taiwan, R.O.C.
[c] National Taipei University of Technology, Taipei 106, Taiwan, R.O.C.
[d] Department of Mechatronic, Energy and Aerospace Engineering, National Defense University, Tahsi, Taoyuan 335, Taiwan, R.O.C.
[e] Advanced Diamond Solutions, Inc., 351 King Street Suite 813, San Francisco, CA 94158, U.S.A.

All conventional pad conditioners dress the CMP pad with discrete diamond grits of random orientation. With the advent of polycrystalline diamond (PCD) dressers, the pad conditioning can be performed with an unprecedented regularity. The result is a total elimination of "killer asperities" that may scratch the soft copper layer or porous dielectric layer. Moreover, the PCD dressers may be constructed to form a blade shaver so pad disruption during conditioning can be minimized. As a result, the original pad polymers can be preserved for polishing wafers. This ultimate diamond disk (UDD) has transformed stick slipping pad destruction to smooth shaving pad construction. UDD is particularly suitable for dressing CMP pads for polishing 22 nm interconnects, in particularly, with the anticipated debut of 450 mm pancake wafers scheduled for 2012.

PCD Cutter

Instead of penetrating the pad surface with a brutal force of diamond grits, advanced diamond disk (ADD) is made by carving a disk of sintered polycrystalline diamond (PCD) layer to form cutting tips of predetermined shape and distribution. The tailor made cutting tips with sharper angle can be formed with higher regularity than conventional diamond disks.

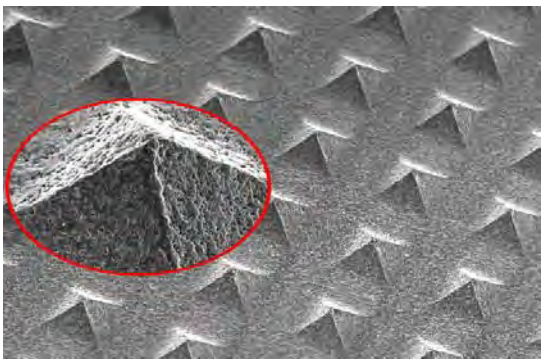


Figure 1. The ADD cutting tips bonded monolithically in PCD matrix material.

Ex. 1016, Page 551

As a result of improved leveling of cutting tips of ADD, the contact points on pad during dressing increases by more than tenfold compared to BDD, the conventional brazed diamond disks.

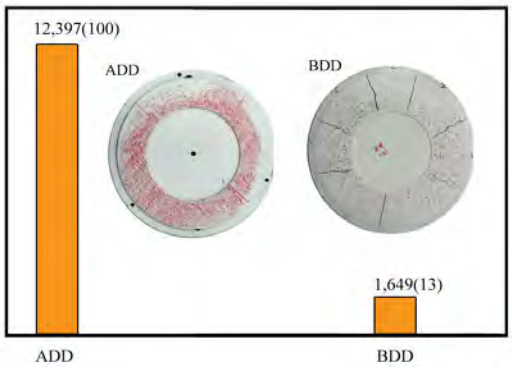


Figure 2. The comparison of working crystals (darkened dots) between ADD and BDD after dressing the pad (IC1000). Both disks have about the same number of cutting tips to begin with the test.

Tender Polishing versus Brutal Polishing

Due to the same depth of cutting, ADD is uniquely capable to form uniform asperities on CMP pads. Moreover, the highly protruded killer asperities are eliminated. The uniform loading of pressed wafer can allow fast removal of IC layer (e.g. copper) at high pressure without causing damages (e.g. scratches) due to the low stress contact.

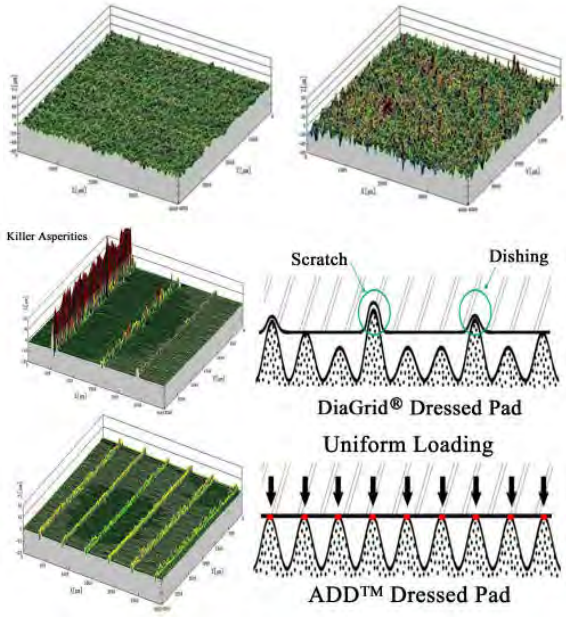


Figure 3. The uniform pad dressing of ADD can avoid the formation of killer asperities that may destruct delicate IC wafers. In the above legend, DiaGrid® refers to BDD pad conditioner.

The CMP with ADD Performance

It has been demonstrated that ADD dressed pad could polish TEOS wafers at a higher removal rate with a better uniformity than that polished with BDD dressed pad.

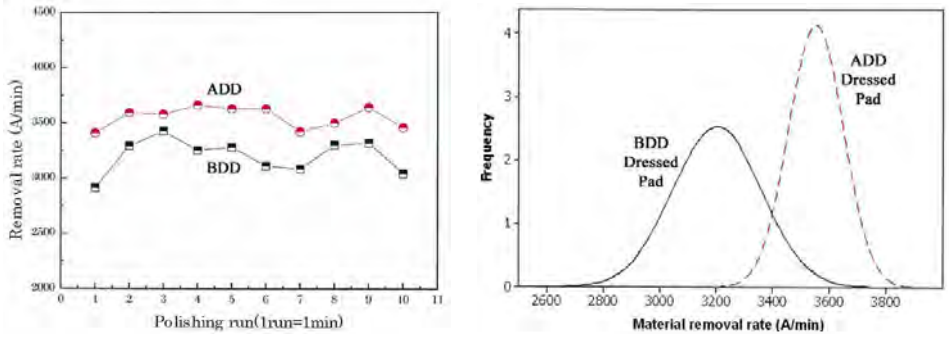


Figure 4. The PCD cutters dressed pad could polish wafer faster with more uniform loading than diamond grits.

The uniform asperities can not only polish wafer with low contact stress, but they are more readily wetted by slurry. As a consequence, the slender asperities can be permeated with abrasive impregnated solution of chemicals so the polishing is performed without dry spots. The abrasion of dry spots would generate both high stress and high temperature; so delicate IC could be damaged.

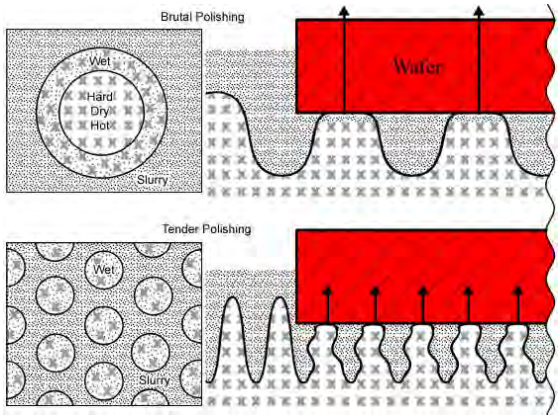


Figure 5. The contrast between brutal polishing with BDD dressed pad and tender polishing with ADD dressed pad.

Soft Landing CMP

In addition to enable high pressure CMP with low stress polishing, ADD's sharp teeth are uniquely capable to cut resilient pads (e.g. Rhom Haas's EV4000) without causing excessive tearing. The resilient pads are designed to reduce local contact stress during polishing. However, due to the presence of bouncing polymers in the matrix, diamond grits cannot cut the pliable pad effectively. In contrast, the rugged PCD cutting tips can break the soft surface and allow the resilient pad to polish wafers consistently.

Ex. 1016, Page 553

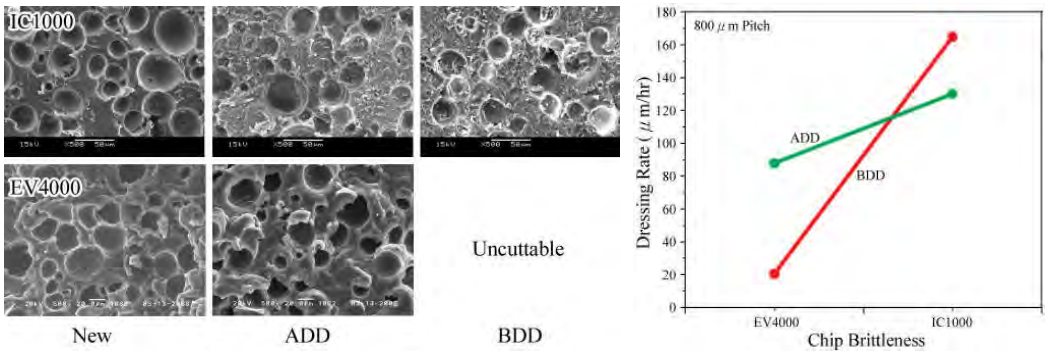


Figure 6. BDD can dress brittle pad materials that may scratch wafer surface. However, ADD can dress resilient polymers to ensure soft landing of delicate IC during CMP.

The Ultimate Diamond Disk (UDD)

Although PCD is shaped on surface to form ADD patterns, there is a better design with PCD sliced to form cutting blades. In this way, the blade can be arranged to make an optimized penetration angle that has not been achievable even with ADD.

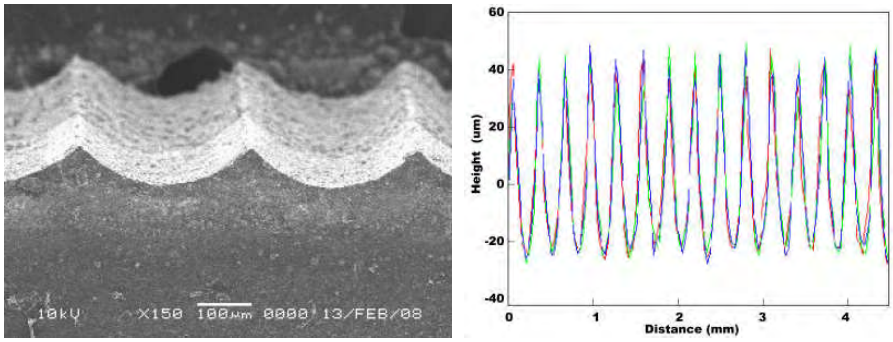


Figure 7. The PCD blade with cutting teeth that can be fully leveled. Note that the cutter tips are leveled within a few microns compared to conventional range of 50 to 100 microns for highest points of discrete diamond grits.

The blades are then arranged in a predetermined pattern and cast in epoxy to form a strange looking diamond disk that is totally out of conventional wisdom.



Figure 8. The UDD with the design of sliding blades. Multiple design features can be accommodated with number of blades and tilt angles.



Figure 9. Additional variations of UDD in blade patterns. The left diagram shows "Chaos Disk", and the right diagram "Sun Disk".

The pad asperities profile formed by UDD is intermediate between ADD and BDD. However, this may be misleading due to the fact that UDD is uniquely capable to cut pad under tension while all other pad conditioners may only drag the pad during cutting. This means that UDD dressed pad surface can retain the most original pad material features of all diamond disks.

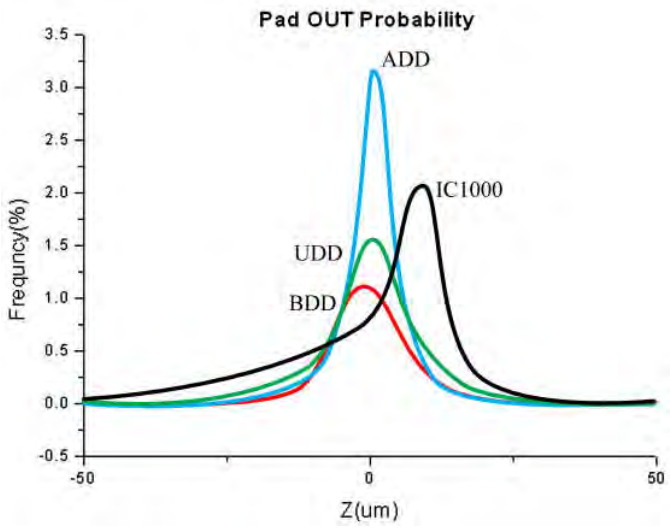


Figure 10. The pad asperities distributions after dressing with ADD, BDD and UDD. Note that the unused pad (IC 1000) was fully leveled on the tips of asperities due the slicing by a sharp steel knife. ADD has the best leveled tips after the pad was dressed with more working crystals. Both UDD & BDD had fewer working crystals so the dressed pad had uneven heights of asperities. In general, the leveling of asperity tips allows uniform and consistent polishing of wafers.

Tensile Shaving Versus Compression Grooving

As UDD is made of arranging cutting blades, many new design concepts previous unachievable can now be entertained. For examples blades with straight edge or with teeth can be bundled to achieve special effects. Straight edges are efficient to scrape clean glazed layer. The cutting tips are best in shredding soft polymers.

Ex. 1016, Page 555

The most important feature of PCD blade is that it can reduce the drag force on the pad and begin to shave polymer. For conventional brittle IC pads, the plastic deformation is not excessive so dull monodiamond grits can still penetrate the polymer. But even so, the polymer is over heated and the chains disrupted. For soft landing pad (e.g. EV4000 of RHEM), not only plastic deformation is large, but also elastic deformation is significant. In this case, monodiamond grits can only deform the pad with minimal penetration. The PCD blade is uniquely capable to avoid the compression of the polymer before penetrating. In fact, the polymer is under tension and it is severed by sharp edge.

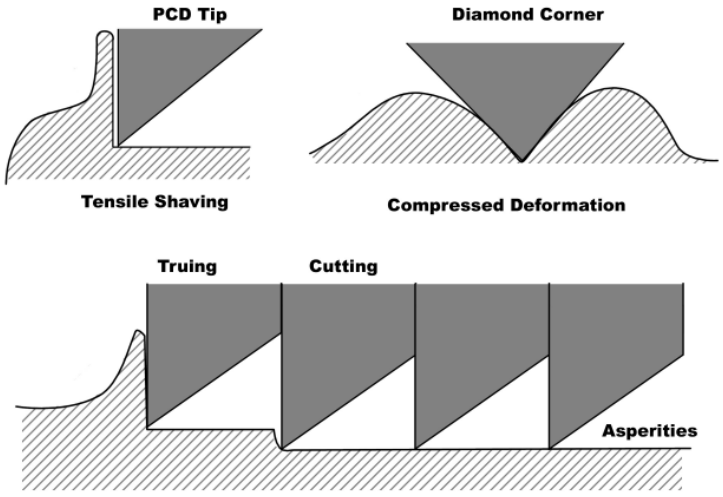


Figure 11. The shaving by tension versus the deforming by compression (top diagram). The blade train that can true the pad before grooving (lower diagram).

Conclusions

PCD dressers are a revolutionary design for CMP pad conditioning. Due to the sculpturing capability, the cutting pyramids can be shaped with high symmetry. The unprecedented tight leveling of the cutting tips can avoid wafer-destructing asperities. Moreover, the PCD can be sliced to make cutting blades. By arranging many blades on a common plane, the ultimate diamond disk can be made. UDD can shave the pad instead of smearing cutting that is prevailing by dressing pad with conventional pad conditioners. The tensile shaving of the pad can preserve the original polymer for polishing IC wafers clean. The following references documented the CMP performance of wafer polishing.

References

1. C. M. Sung, S. C. Hu, W. Huang, C. S. Chou, C. C. Chou, and Y. L. Pai, “The Design Features of Ultimate Diamond Disks (UDD): W CMP with In-Situ Dressing of Metal Free Diamond Disks” (2008).
2. C. M. Sung, C. S. Chou, C. C. Chou, Y. L. Pai, W. Huang, S. C. Hu, E. Chou, L. Liao, T. C. Hu, H. Y. Lin, A. Wang, H. C. Lee, and M. Sung, ”CMP Pad Dressing by Avoiding Killer Diamonds” (2008).

Ex. 1016, Page 556

ECS Transactions, 18 (1) 529-534 (2009)
10.1149/1.3096497 © The Electrochemical Society

Development of CMP Conditioner for Sub-50 nm Technology Nodes

Taewook Hwang[a], Rama Vedantham[a], Simon Zhang[b]

[a] Saint-Gobain Abrasives, One New Bond Street, Worcester, Massachusetts 01615, USA
[b] Saint-Gobain Abrasives（Shanghai）, No.198 Bei Dou RD, Minghang Development Zone,Shanghai,China

Several key conditioner design parameters including diamond size, shape, and concentration were selected and optimized to tune wafer removal rate and pad cut rate for various types of CMP pads. Among various diamond types, two types of diamonds were selected for conditioner applications and numerous conditioners were manufactured to generate pad surface roughness on a regular polyurethane porous pad, hard porous pad, and hard solid pad. Furthermore pad cut rate stability curves and wafer removal rate were generated to correlate conditioner design with pad surface texture and wafer removal rate and defects. Experimental results revealed the effect of conditioner design on pad surface roughness was significant and wafer removal rate varied significantly depending on a specific set of conditioner and pad.

Introduction

As IC technology continues downsizing to 45 nm and 32 nm feature sizes, tight defect control becomes the most single important issue for various cmp processes. This requirement intensifies the challenges faced by suppliers of pads, slurries, and conditioners, and also drives collaboration between these suppliers to optimize consumable sets. During the conditioning process, it is not sufficient to simply maintain process stability by conditioning the glazed pad surface. In addition, the conditioner is also responsible for generating pad texture for various types of different pads. Development of advanced conditioners requires an investigation of interaction mechanisms among conditioner, pad, and wafer [1-2]. Recently various types of CMP pads have been actively introduced to IC Fabs to improve cost of ownership and reduce sub-micro scratches. Microstructures of CMP pads are quite different from supplier to supplier with some porous pads having either opened or closed cell micro structures and others having non porous structures. Non porous type pads usually require additional process to make holes by either machining or laser processes to enhance slurry delivery into the polishing zone. With various types of commercially available CMP pads, optimization and customization of conditioner design for each type of CMP pad should be made to maximize pad and conditioner performance. The objective of the present paper is to develop optimized CMP pad conditioners for sub-50 nm technology nodes by investigating effects of key CMP conditioner design parameters on various CMP pads.

Diamond Grit Cutting Test and Selection

The crystal shape influences many diamond attributes such as crystal toughness and sharpness. Desirable pad cut rate can be obtained by proper selection of diamond shapes and their relative composition. Prior to using a regular 4 inch diameter CMP conditioner

Ex. 1016, Page 557

to measure pad cut rate, an experimental CMP conditioner having only a limited known number of diamonds on its top surface was made. Normal and tangential cutting forces were measured while the experimental conditioner cut on a polyurethane CMP pad. Three types of diamonds, designated as "Very Sharp", "Sharp" and "Dull", were tested with an experimental conditioner. Then the tangential and normal force per grit for each diamond type were obtained by dividing the total measured tangential and normal cutting force by the known number of diamonds. The normal force per grit in Fig. 1 ranged from 0.13 Newton for "Dull" grit to 0.26 Newton for "Very Sharp" grit. After a cutting test with three diamond types, many other diamond types were identified and the pad cut rate of each diamond type was measured with a regular conditioner manufactured with the same design parameters except diamond type. For an example in Fig. 2, the pad cut rate significantly varied depending on a type of diamond and the pad cut rate of Type-E was more than 20 times higher than Type-A.

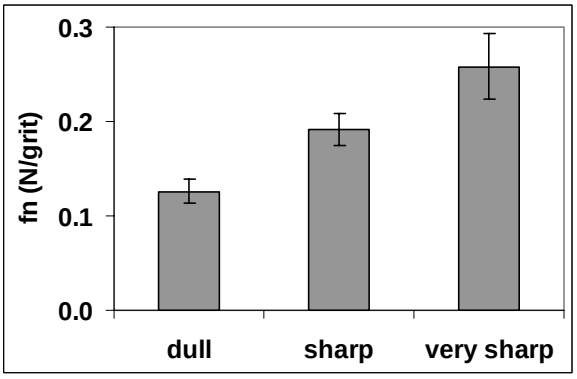


Fig. 1. Normal cutting force per grit (fn) for three different diamond types.

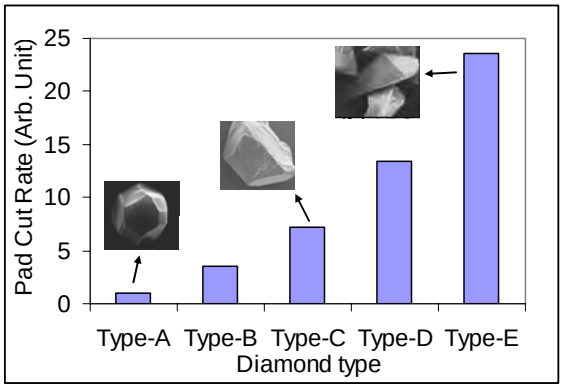


Fig. 2. Pad cut rate of various diamonds.

Pad Surface Analysis and Conditioner Optimization

Pad cut rate and arithmetic mean surface roughness, R_a, were generated with one specific conditioner design (SGA-C in Fig. 3) on four different types of commercially available CMP pads designated as: "Regular Porous", "Soft Solid", "Hard Porous", and "Hard Solid" in Table 1. The pad cut rate of Soft Solid pad was 36 times higher than the Hard Solid pad. A finer surface finish was obtained with both solid pads and the R_a of Hard Porous pad was slightly lower than the Regular Porous pad. This result raises the issue of whether optimization of conditioner design and conditioning parameters for each specific pad is needed to obtain desirable pad cut rate and pad texture.

Table 1 Comparison of pad cut rate and surface roughness for four types of pads with a SGA-C conditioner in Fig. 3

Pad Type	Regular Porous	Soft Solid	Hard Porous	Hard Solid
Pad Cut Rate (Arb. Unit)	4.0	72.0	2.2	1.9
R_a (um)	2.8	1.7	2.5	1.5

Ex. 1016, Page 558

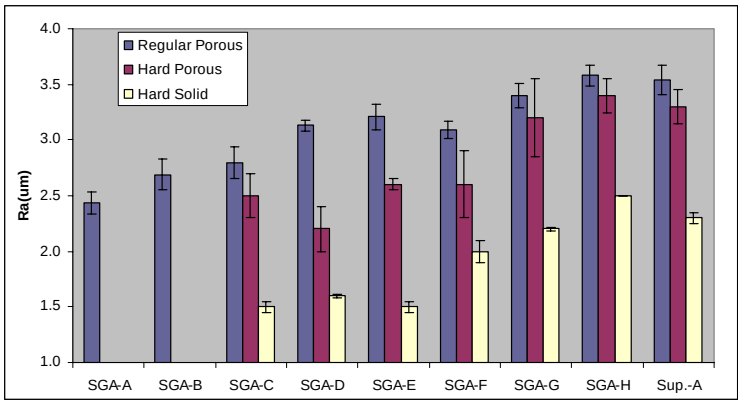


Fig. 3. Arithmetic mean surface roughness R_a for three types of CMP pads conditioned with various conditioner designs.

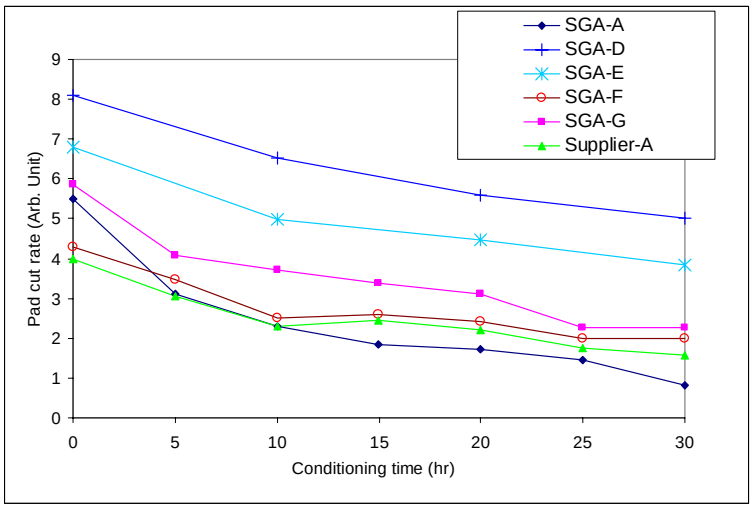


Fig. 4. Pad cut rate curves versus conditioning time for various conditioners.

Several key conditioner design parameters were investigated to determine their effects on surface roughness and wafer removal rate [3, 4]. A total of 8 different specifications, designated as SGA-A~H, were selected after initial screening test. One conditioner from another supplier, designated as "Supplier-A", was included for benchmark purpose. Three types of pads, Regular Porous, Hard Porous, and Hard Solid were conditioned for pad surface roughness evaluation with the given nine conditioners. Fig. 3 shows variations of R_a versus each conditioner design for three types of pads. The values of R_a on Hard Solid pad were much lower than those on other two porous type pads. The R_a values also significantly varied depending on each conditioner design. The surface roughness values in Fig. 3 were obtained with each new conditioner after break-in. In order to investigate time dependent behaviour of pad surface finish and pad cut rate with an increased conditioning time, five SGA designs and one Supplier-A conditioner were conditioned up to 30 hours on a Regular Porous pad. Then pad cut rate was measured every 5 or 10 hours. As expected, pad cut rate decreased with conditioning time as shown in Fig. 4, and SGA-F and Supplier-A followed a similar curve while the pad cut rates of SGA-D&E after 30 hours of conditioning were still above 4. Fig. 5 shows R_a values on a Regular Pad, conditioned with the five conditioners before and after 30 hours conditioning. There was no significant variation of pad surface roughness with the five

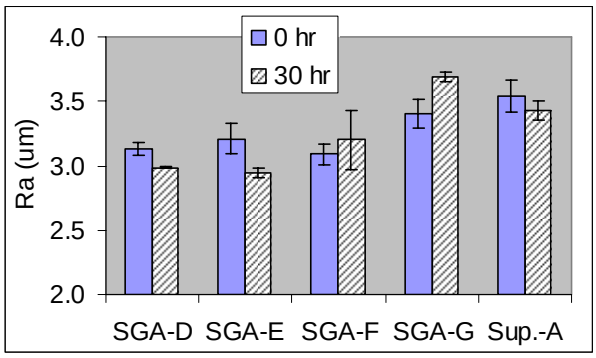


Fig. 5. Pad surface roughness R_a before and after 30 hours of conditioning.

conditioners although pad cut rate of each conditioner decreased significantly from its initial value after 30 hours.

Wafer Data and Discussions

In order to see whether wafer removal rate varies depending on each pad and conditioner set, a polishing test was conducted with five selected conditioners (SGA-A, E, F, G, Supplier-A) and two different pads (Regular Porous pad and Hard Solid pad) under typical oxide CMP conditions with in-situ conditioning and 5 lbf down force using a commercial fumed silica oxide slurry. The normalized wafer removal rates plotted versus each conditioner design for two types of pads are shown in Fig. 6. For Hard Solid pad, the wafer removal rate of SGA-A is about 20% higher than that of Supplier-A conditioner. Effect of conditioner design on wafer removal rate for Regular Porous pad is less significant than Hard Solid pad. Overall removal rates with Regular Porous pad are higher than that with Hard Solid pad. Removal rate data are further plotted versus corresponding surface roughness values in Fig. 7. From the results in Fig. 10, wafer removal rate decreases as surface roughness increases for both types of pads. It can be inferred from strong correlations between surface roughness and wafer removal rate that the generation of fine pad surface would be desirable, leading to higher wafer removal rate and presumably improved defects.

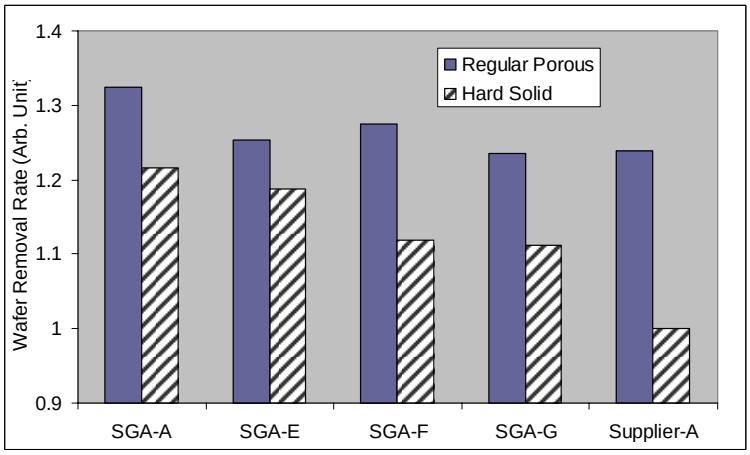


Fig. 6. Wafer removal rate versus each conditioner design for Regular Porous pad and Hard Solid pad.

Further experimental validations conducted at several high volume production Fabs also confirm that a finer pad surface, conditioned by an optimized conditioner, increases wafer removal rate and reduces defects [3, 4]. Generation of fine pad texture becomes even more critical for sub-50 nm technology nodes to achieve good planarity with minimal dishing. Fig. 8 illustrates planarity data of post-CMP oxide trench dishing depth obtained from 300 mm production patterned wafers polished with Regular Porous pad. The dishing depth with SGA-A conditioner is less than one third of that with Supplier-B conditioner.

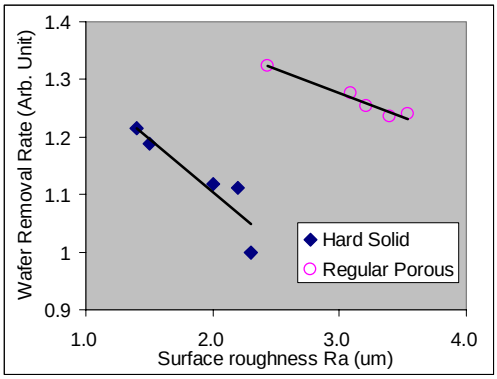


Fig. 7. Correlation between wafer removal rate and pad surface roughness.

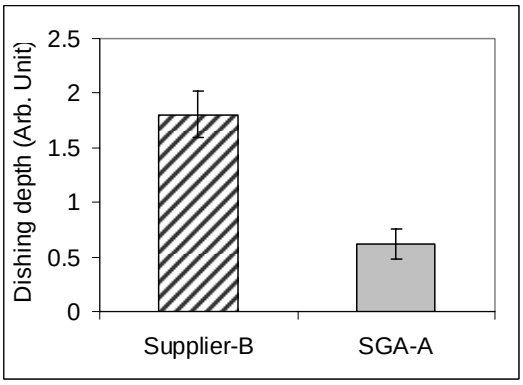


Fig. 8. Average dishing depth of oxide trench with two types of conditioners obtained from Regular Porous pad.

CONCLUSIONS

Conditioner design optimizations followed by experimental validations have been conducted for various types of CMP pads. It is found that pad surface roughness R_a significantly varies depending on a specific set of pad and conditioner. Consequently, wafer removal rate is significantly influenced by each type of pad and conditioner set. For Hard Solid pad, wafer removal rate increases up to 20% with an optimized conditioner design whereas for Regular Porous pad wafer removal rate increases up to 7%. This variation in wafer removal rate is found to have a strong correlation with corresponding pad surface roughness value. The wafer removal rate increases as pad surface roughness decreases. Experimental validations in high volume production Fabs confirm that fine pad surface, conditioned by an optimized design, delivers increased wafer removal rate and reduced defect.

Acknowledgments

The authors would like to thank T. Puthanangady, T. Wu, C. Dinh-Ngoc for valuable discussions and contributions.

References

1. T. W. Hwang, G. Baldoni, A. Tanikella, T. Puthanangady, “Advanced Pad Conditioner Design for Oxide/Metal CMP”, Presented at 2nd PacRim ICPT(2005), published in Transactions on Electrical and Electronic Materials, Vol. 7, No. 2, p.62, April 2006

2. A. Scott Lawing, " Polish rate, pad surface morphology and pad conditioning in oxide chemical mechanical polishing", The Electrochemical Soc. Proc., Vol. PV 2002-1, p. 46, 2002
3. J. C Yang, T. S. Kim, T. W. Hwang, T. Puthanangady, "Optimization of CMP conditioner design for non-porous pad and its application to manufacturing process", Presented at the 13th International Symposium on Chemical-Mechanical Planarization (CAMP), Lake Placid, NY, USA, August 10-13, 2008
4. T.W. Hwang, G. Baldoni, R. Vedantham, T. Puthanangady, "Optimized and Customized CMP Conditioner Design for Next Generation Oxide/Metal CMP", International Conference on Planarization/CMP Technology (ICPT), Dresden, Germany, p. 397, October 25-27, 2007

ECS Transactions, 18 (1) 535-540 (2009)
10.1149/1.3096498 © The Electrochemical Society

The Design Features of Ultimate Diamond Disks (UDD): W CMP with In-Situ Dressing of Metal Free Diamond Disks

James C. Sung[a,b,c], Shao-Chung Hu[a], Wei Huang[a], Cheng-Shiang Chou[a]
Chih-Chung Chou[a], and Yang-Liang Pai[a]

[a] KINIK Company, 64, Chung-San Rd., Ying-Kuo, Taipei Hsien 239, Taiwan, R.O.C.
[b] National Taiwan University, Taipei 106, Taiwan, R.O.C.
[c] National Taipei University of Technology, Taipei 106, Taiwan, R.O.C.

All pad conditioners used for CMP today are made by attaching discrete diamond grits on a flat substrate. Polycrystalline diamond (PCD) was wire-EDM cut to form blades with different designs of cutting tips. These blades were assembled to form the so-called ultimate diamond disk (UDD). The pad asperities formed with UDD were compared with conventional diamond disks. It was confirmed that UDD dressed pad asperities were more uniform than that for polishing IC wafers today. The UDD dressed pads were also used to polish tungsten (W) IC layer deposited on 300 mm wafers. As the surface of PCD contains only merged diamond grains that are corrosion immune, UDD was dressing in-situ, by immersing in the acid slurry. The wafer uniformity so polished was comparable to that of ex-situ CMP pad dressing with metallized diamond disks. The in-situ dressing can boost the CMP throughput by deglazing the pad in real time while the wafer is being polished. After 15 hours, the wafer profiles for this in-situ dressing were sustained without decay.

Erratic Dressing of CMP Pads by Diamond Grit Pad Conditioners

All pad conditioners used today are made by attaching individual diamond grits on flat substrate. Due to the variations of diamond sizes and shapes, the pad dressed contains random asperities. Consequently, CMP performed by pressing IC wafers on the chaotic pad landscape may not be controllable with undesirable profiles and inevitable defectivity.

Ex. 1016, Page 563

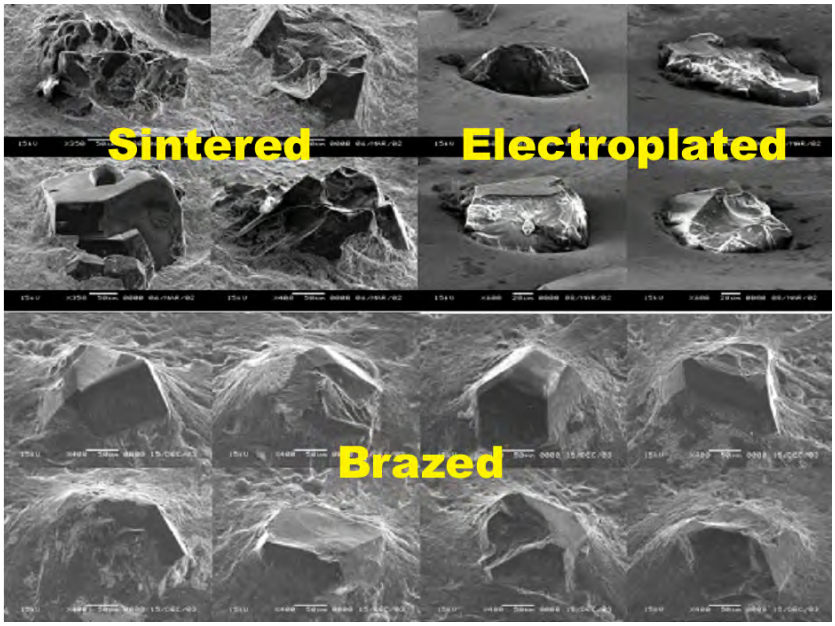


Figure 1. The irregular appearance of metal attached diamond grits on CMP pad conditioners.

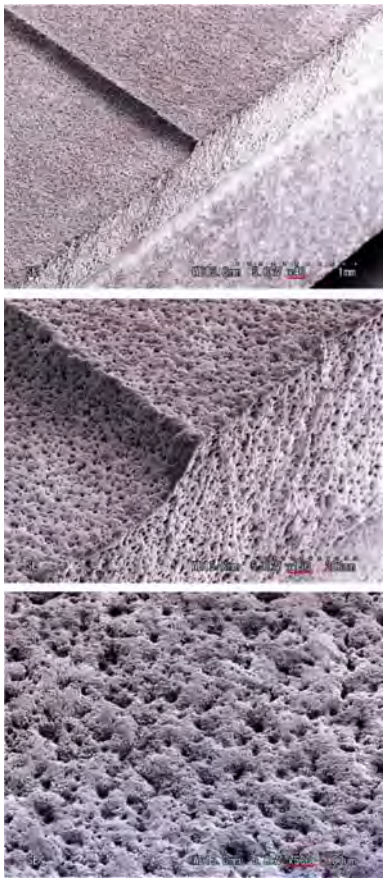


Figure 2. The PCD surface cut by wire-EDM revealed interconnected diamond networks with trapped cobalt lost by evaporation during sparking with electricity. Note that diamond grains are mutually bonded by diamond bridges, so the plucking out of individual particles during CMP dressing can be avoided.

The Ultimate Diamond Disk with PCD Blades

PCD blanks were sliced to form seriated blades. These blades were then assembled by casting in a strong epoxy. The ultimate diamond disks (UDD) so formed contained cutting tips with a predetermined shape and distribution. The cutter tip leveling was about one order of magnitude better than conventional pad conditioners.

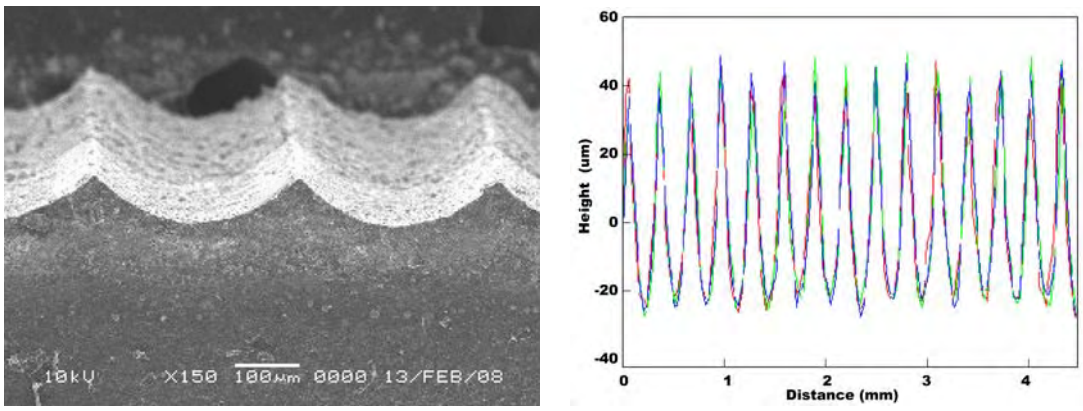


Figure 3. The PCD blade with cutting teeth leveled within a few microns. This range was much tighter than that (50 to 100 microns) of conventional diamond disks with discrete diamond grits.

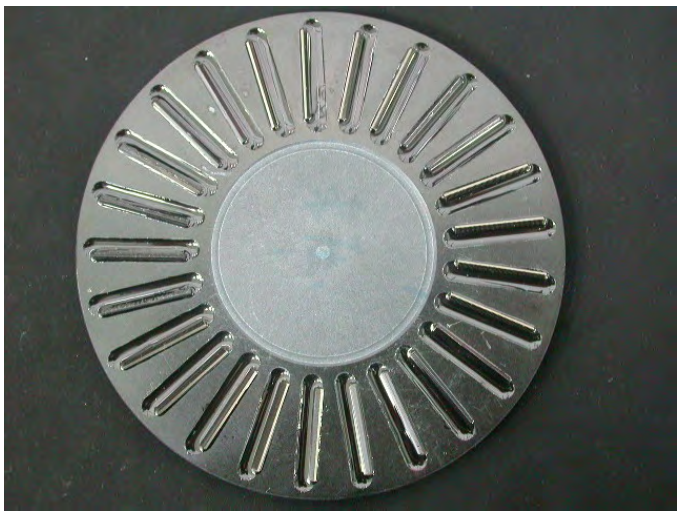


Figure 4. The outer appearance of UDD with 26 cutting blades that were used in dressing CMP pads for polishing tungsten layers on IC wafers.

The Control of Pad Dressing Rate

Conventional diamond disks are measured by a primary feature, the cuttability of pads as measured by dressing rate (e.g. 100 microns per hour). Due to the uncontrolled diamond sizes and shapes, the cuttability is typically tuned by selection of diamond shapes. This can be done by spreading diamond particles across an inclined table that vibrates asymmetrically as to move irregularly shaped diamond to the far end. In general, the more regular the diamond shape, the lower the dressing rate. To achieve an optimized polishing rate of the wafer during CMP, the glazed layer of the pad must be

Ex. 1016, Page 565

removed periodically with the restoration of sharp asperities. However, over dressing is not only wasting pads, but also risking wafers. The fast cutting of pads is achieved by sharp diamond corners that are vulnerable to break. The broken diamond shards can cause severe arc scratches on delicate IC wafers.

Instead of using vulnerable sharp edges, UDD's cuttability is determined by the angle of the cutting teeth and their pitch. Consequently, the polishing rate of the wafer can be tuned by dialing a dressing rate of the pad. Moreover, the UDD blades can be designed in such a way that no cutter can penetrate the pad too deep as to cause breakage.

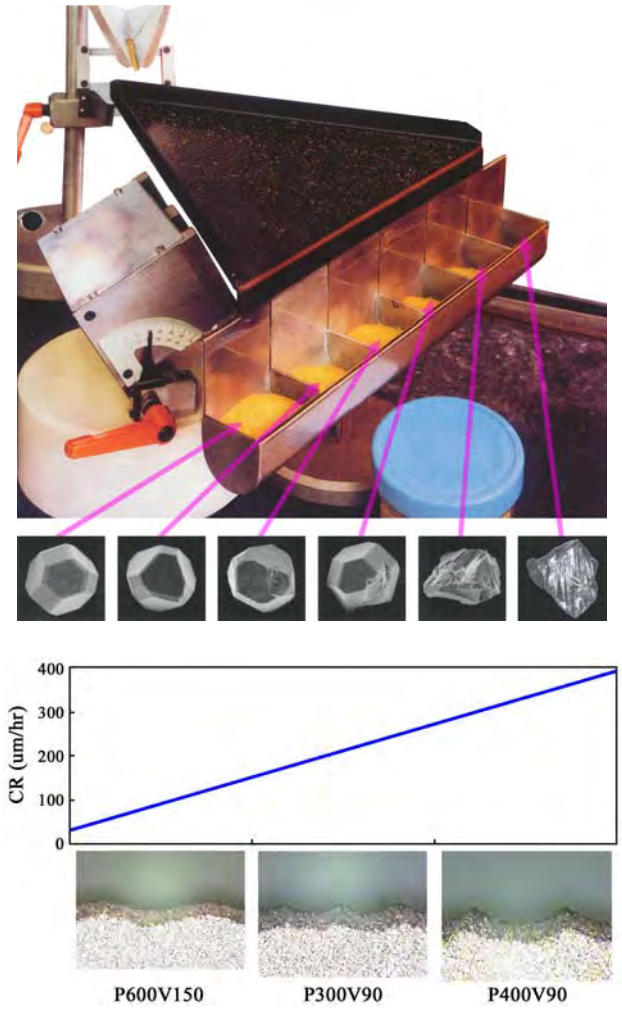


Figure 5. The statistical control of pad dressing rate by shaping the diamond grit and the engineering control by making the PCD teeth. In the above diagram, P denotes pitch and N for angle for PCD cutting blades.

Several different designs of UDD blades were made to dress CMP pads. All of them are capable to form asperities more uniform than conventional metallized disks. Moreover, the sharper the teeth, the faster the dressing rate of the pad. However, the leveling of the asperity tips depends on the flatness of the PCD blade. The straight blades in essence form a tueing diamond disk (TDD). TDD dressed pads possess the least variation of the asperity tips and with the longest life. However, the compromise is to polish wafers slower due to the increased number of contacting asperities.

Ex. 1016, Page 566

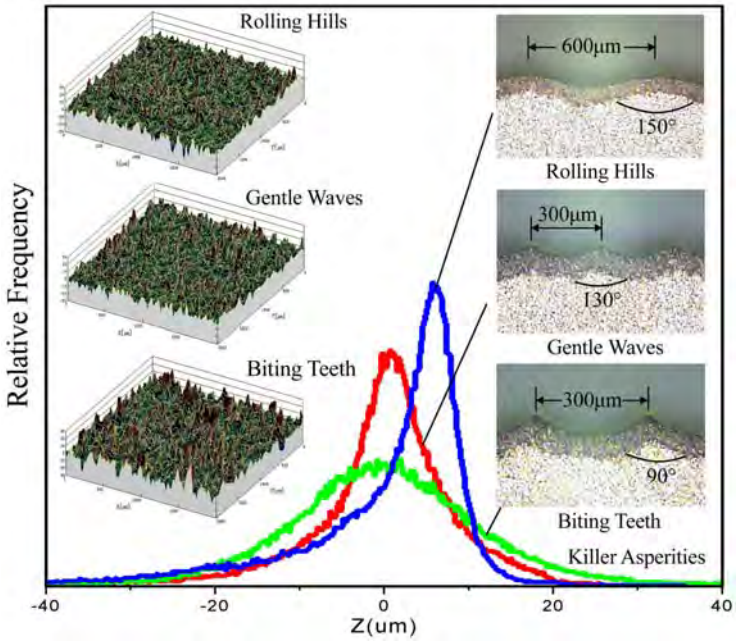


Figure 6. The pad profiles dressed by UDDs with different blade geometry. Note that the asperities profile can be altered by changing the teeth sharpness.

The above figure demonstrated that the gentler the PCD profile, the more effective to deglaze the pad, the tips of pad asperities are also better leveled. Although the pad life is extended by cutting less, the contact area between wafer and pad during polishing may increase as to cause the decrease of polishing rate. Hence, the pad life and wafer-polishing rate can be compromised to optimize the CMP performance.

In addition to make cutting blades, the entire PCD blank may be shaped by wire-EDM to form cutting pyramids that may be distributed on the annular surface of the pad conditioner. This design was known as advanced diamond disks (ADD). The comparison of dressing rates among the three PCD cutting designs is shown below.

TABLE I. The Dressing Rate Comparison.

Design	Cutter Tip Angle	Micron Per Hour
ADD	90	100
UDD	150	90
TDD	180	20

TDD with low dressing rate of the pad but with high contact edge length is ideal for deglazing of soft pads, such as politex for barrier polishing (P3). The soft pad is made of non-woven fibers that cannot be effectively cut with dull diamond of large sizes. Instead, a cutting blade with small diamond crystals that are tightly packed is needed. The TDD blades contain the tightest diamond grains possible with individual cutting points only ten microns in size. Such small and sharp teeth are efficient in dressing soft pads.

The CMP Performance of UDD by In-Situ Dressing of WCMP

The current diamond disks with chaotic cutting tips cannot be optimized for the current CMP manufacture. UDD with control of cutter geometry and blade separation allow the

Ex. 1016, Page 567

finer tuning of the design to control the CMP performance to achieve desirable features.

UDD was tested by replacing metallized diamond disk (brazed diamond disk or BDD, i.e., DiaGrid® pad conditioner) for dressing pads for polishing tungsten layer on 300 mm wafers. WCMP is normally conducted with ex-situ dressing due to the presence of metal layer on the pad conditioner that may be dissolved fast in acidic slurry. A typical metallized diamond disk cannot survive the in-situ dressing for more than a couple of hours. After then, diamond grits begin to dislodge and the expensive wafer may be scratched badly. In contrast, UDD was used in-situ along with the polishing action of the IC wafer. The result indicated that UDD could be a drop in replacement of metallized diamond disks, but with the potential benefit to gain higher throughput for the CMP production.

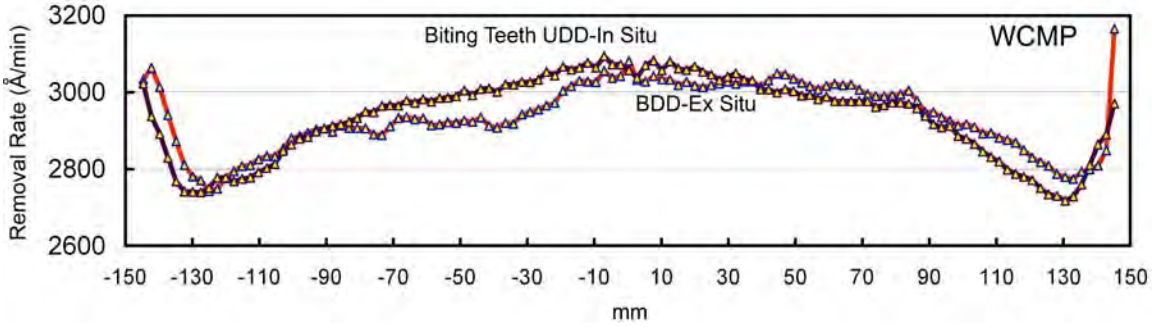


Figure 7. The wafer profile for WCMP with in-situ dressing of UDD was comparable to that with ex-situ dressing of BDD.

Additional endurance tests of UDD in-situ dressing were conducted for polishing W layer on 300 mm wafers. The result demonstrated that UDD maintained the polishing rate after more than 15 hours of in-situ dressing. In fact, the polishing rate was even increased with time. This result confirmed that UDD could be used to sustain high productivity of CMP by minimizing the machine down time for switching pad conditioners.

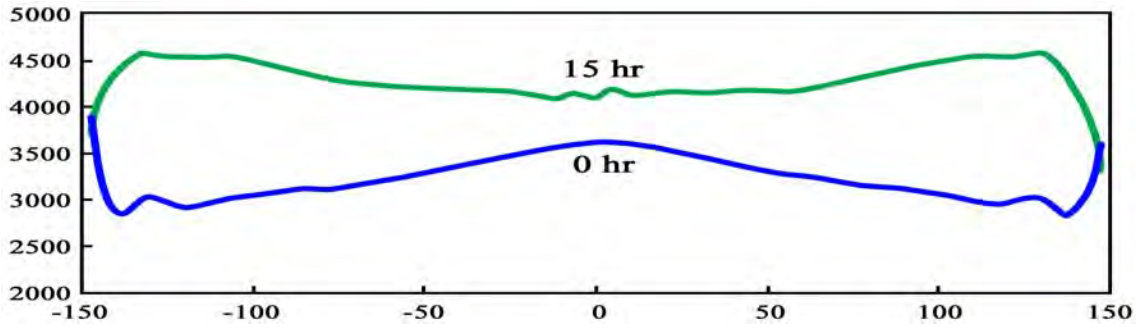


Figure 8. The wafer profile of W layer after in-situ dressing of pad for 15 hours.

Conclusions

A revolutionary design of pad conditioners by arranging PCD blades on the diamond disk could boost the CMP throughput by in-situ dressing the pad in acidic slurry. The metal free diamond disk may also avoid contamination that is inevitable with conventional pad conditioners. CMP performance can be optimized by using UDD with different designs.

Ex. 1016, Page 568

ECS Transactions, 18 (1) 541-546 (2009)
10.1149/1.3096499 © The Electrochemical Society

Slurry Development Challenges and Solutions for Advanced Node Copper CMP

Dr. Deepak Mahulikar
Sr. Technical Director
Planar Solutions LLC USA
Deepak_Mahulikar@fujifilm-ffem.com

The R&D of Cu CMP started in the late 1980s but was fully implemented in the HVM in the late 1990s and early 2000.Cu CMP was fully implemented for the 130 nm nodes primarily in the Logic devices. In the decade following the Cu CMP introduction, it has grown and changed in significant ways. As of now Cu CMP is a main stay for most BEOL integration schemes and being applied to Memory, 3D and MEMS devices.

The success of Cu CMP was due to many enabling events such as Ta barrier development, AMAT Mirra tool design and low selectivity barrier process. As we move from 130 nm nodes into more advanced nodes additional challenges have unfolded. Performance requirements have become quite stringent while CoO reduction pressures are mounting. While the 130 nm process required polishing of Cu, Ta and TEOS at high DF , moderate topography and moderate defects, for 32 nm and beyond polishing of multiple complex materials (like LK, ULK, ALD barriers, MHM, ARC and Copper) at very low DF with very low topography and defects is essential.

These challenging needs for the advanced nodes require novel slurry development solutions. Both the Copper and Barrier layer slurries have changed over the decade. Abrasives have changed from Alumina and fumed silica to colloidal silica and coated particles. Significant change in chemistry design is required to address low DF polishing.

Very low defects is a key need and challenge for the advanced nodes. The defects are impacted not only by slurry chemistry, abrasives and LPCs but also by Trace Metals, pad interactions, conditioner interactions and post CMP processes.

Other challenges include stable Concentrate development, and slurry design for lower CoO. This paper will outline those challenges and new novel approaches to resolve some of those issues

Cu Integration Past, present and future

Cu, as an interconnect material, was fully integrated primarily in the 130 nm node in the late 1990s to early 2000. The success of Cu integration was driven by many factors including reliable materials and deposition, Ta/TaN as an effective barrier for Cu diffusion and of course Cu CMP. The 130 nm integration stack (see Fig.1) was very simple. As we progressed to 90, 65 and 45 nm integration materials and processes became complex and the future (32, 22 nm) looks even more complex. These comparisons are shown in Fig.1.

Ex. 1016, Page 569

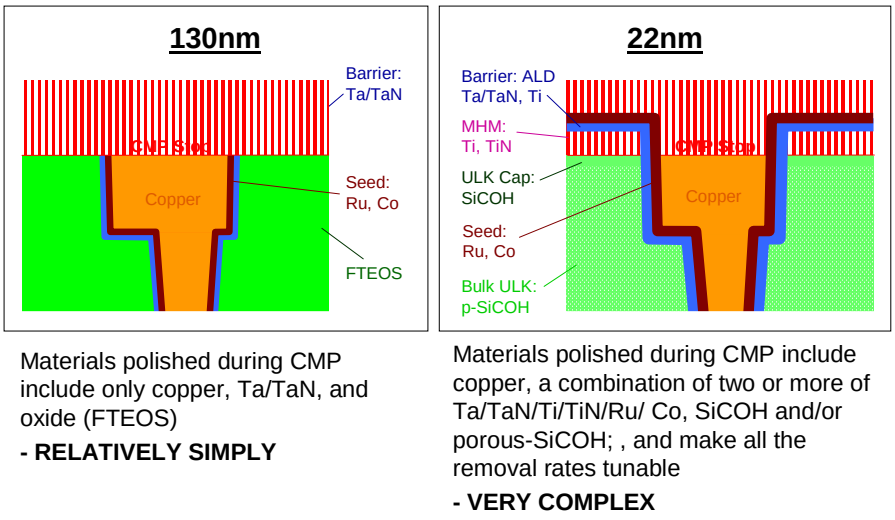


Figure 1

The complexity in the 22 nm integration presents many issues as shown in the Figure 2

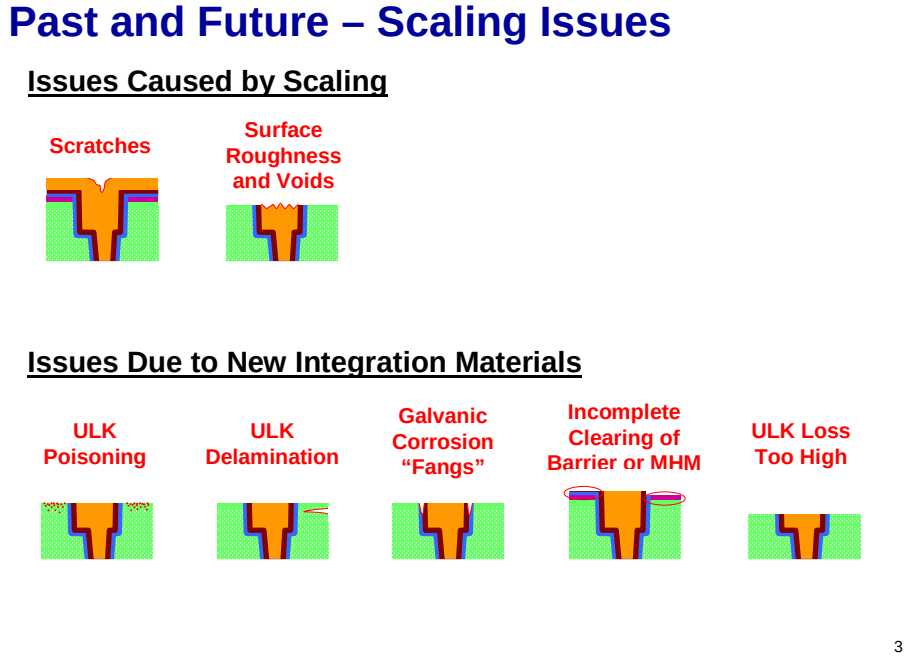


Figure 2

Differences are outlined in the Table below

Table 1 : Comparisons 130 versus 22 nm nodes

130 nm	22 nm projections
Short list of materials needed to polish for Cu CMP : TEOS, Ta/TaN, Cu	**Material list long and exotic ULK, LK, Ti, Ta/TaN, ALD, Co, Ru, TEOS, ARC, Cu, Cu alloys**
Relaxed performance needs **➢ 500A Topography** **➢ >100 defects OK** **➢ High DF process**	**Extraordinary performance needs** **< 200A Topography** **< 10 defects** **Excellent Planarization Efficiency** **High removal rates @ low DF**
CoO modest	**Low CoO**
200 mm process	**300 mm process**
Throughput < 30W/h	**Throughput > 50 W/h**

Obviously the consumable needs for the Future are different and challenging.

Table 2 Consumables Then and Now

130 nm	45 nm
Hard pads for both Cu and barrier layer CMP	**Multiple type Hard pads for Cu Polish**
Al2O3 based acidic Cu slurry	**Multiple type Soft pads for barrier layer polish**
Fumed silica based low selective barrier layer slurry	**Colloidal silica or designer particle based Cu and barrier layer slurries with varying pH**
Acidic Post CMP and pad cleaners	**Acidic or Alkaline post CMP cleaners**
Standard pad conditioners	**Optimized pad conditioners**
	Customized consumables

The Cu CMP slurry needs are outlined below

- Customized slurries
- Very high performance at much lower CoO
- High Purity
- Excellent product Lot to Lot Stability
- Extended shelf life
- Local Manufacturing
- Compatibility with other consumables AND multiple tool platforms

While 130 nm node used standardized slurries like Planar Solutions Cu10K-2, 32 nm and beyond will need custom slurries. That is because integration schemes are not standardized. Thus one customer may need a direct ULK polish, another may not.

The 32 nm Integration will require exceptional functional performance as listed below

Ex. 1016, Page 571

Table 3 32 nm Performance Needs

Cu slurry	Barrier layer slurry
Cu removal rates 8K/min @ 2-3 psi, >3K/min @1-2 psi, < 500A topography	**Excellent topography correction , final topography < 200 A**
Excellent Planarization Efficiency	**Customized Cu, ILD, barrier layer polish removal rates at very low DF**
Very low scratches, corrosion or residue defects	**Very low defects : corrosion, scratches, Fangs**
Long Pad Life and no staining	**Long pad life**
Stable removal rates during shelf life	**Stable removal rates during shelf life**
Highly selective to barrier layer	**Ultra clean Post CMP surface**
Same slurry for platen 1 and 2	

This performance is needed at much lower Cost of Ownership, which is very challenging for slurry designers. The CoO for slurries is driven by many factors. The needs for high purity, high removal rates at lower DF, and very low defects drive the costs higher.

One of the key change affecting the Semiconductor industry in general is that the majority of business has shifted to Asia. US based slurry makers are therefore faced with decisions of very high freight costs or major investments for Local Mfg.

Additional challenges for slurry makers include slurries compatible with many different pads and tool platforms. Fabs are also advocating “Green” environmental strategy. Thus slurries need to use friendly chemicals. To achieve the high performance listed above at very low CoO, this becomes a significant challenge. We also cannot forget the competitive landscape. There are more than a dozen slurry makers and more keep coming. Implementation and honoring of IP also presents barriers.

Novel and innovative solution approaches to the challenges

As big and complex the challenges are, the slurry developers are addressing them with novel solutions. The very high performance needs are addressed by innovative particle and chemistry design. Much lower defects are achieved by switching to soft colloidal silica or coated particles from Al2O3 or fumed silica particles. Particle surfaces are also being modified to achieve higher removal rates and better stability. Very effective additives have been discovered that reduce topography and increase removal rates. Example is shown in the Table below

Ex. 1016, Page 572

Table 4 : Solution approaches

	130 nm standard Step 1 slurry	130-90 nm changed Cu slurry 9041	45 and below bulk and SL Cu slurry 9044
Particle	Al2O3	Colloidal silica	Colloidal silica
pH	Acidic	Neutral	Neutral
Chemistry	Inorganic acids, oxidizers and CI	Rate enhancers, CIs, oxidizer, organic acid	Rate enhancer, dishing reducers, CIs, novel additive
Performance	Bulk slurry primarily ➢ 5K A/min Cu rate ➢ 200 Selectivity ➢ 1000 A dishing ➢ 1000 defects	Bulk and SL slurry >9K A/min Cu rate >500 Selectivity ~1000 A dishing < 100 defects	Bulk and SL slurry >8K A/min Cu rate >500 selectivity <500 A dishing < 100 defects
Price	100	< 50	< 50

CoO reduction requires addressing materials and process costs. High purity particles and chemistry are expensive. So attempts are made to reduce % solids and make concentrates. Mfg process steps are optimized (filtration for example). Tool flow rates are reduced. Local mfg is being addressed.

Today's slurries therefore have a different look than the past. Trends include concentrates, colloidal silica or coated colloidal silica particles, lower % solids, higher chemistry, and novel chemistry. In addition the number of slurries has gone up dramatically. As Cu integration becomes more complex, slurry development will have to rise to the challenge with even more innovative solutions.

Acknowledgements

Help from Bin Hu, Richard Wen,S. Mohseni and Keven Cline is appreciated. Author thanks Planar Solutions LLC for permission to publish this paper.

References

1 Mahulikar, Wen and Hu, " Slurry design for CMP of Cu-Low K layers on semiconductor wafers" Specialty Chemicals Magazine, March 2005, V.25 N. 2

2 DeNardis et al "Investigating the effects of diluting solutions and trace metal contamination on aggregation characteristics of silica based CMP slurries" MRS Spring Meeting Proceedings, March 2005

3 Miller et al "Influence of slurry chemistry on Cu polish defects" Lake Placid CMP symposium August 2003

4 Gramm, Mahulikar and Adkins, "Effect of organics on the properties and performance of CMP slurries" Lake Placid CMP conference proceedings August 2002

5 Stober et al, J. Colloidal Interface Science, **26** (82) 1968

6 Wen et al " Electrochemical properties of Corrosion behavior of Cu barrier slurries" Lake Placid CMP conference proceedings August 2005

7 Mahulikar et al " Effect of impurities on the CMP performance of colloidal silica and fumed silica based slurries " in PacRim ICPT proceedings 2005

8 Hu et al " High removal rate Cu slurry development for TSV application" in ICPT Taiwan 2008

ECS Transactions, 18 (1) 547-552 (2009)
10.1149/1.3096500 © The Electrochemical Society

Application of Surfactant in Si CMP Processing

B.M. Tan, X.H. Niu, P. Liang, T. Wu, Y. L. Liu
Institute of Microelectronics, Hebei University of Technology, Tianjin, 300130,China

With the rapidly developing of ULSI fabrication technique to high integration and low cost, the key technology is how to ensure the higher flattening, lower roughness and lower cost. The chemical mechanical polishing (CMP) method is the best one for global planarization. In this paper, the CMP mechanism process of silicon wafer was analyzed, and it was pointed that chemical reaction was the kinetics control process. The influence of surfactant on silicon surface quality during CMP process was also discussed, and the contrast experiments were processed. SiO_2 was selected as abrasive, which particle size was 40nm. And the polishing temperature was 25℃. At the end of silicon CMP, the water polishing was processed using de-ionized water with surfactant, and the optimal surface was gotten, and the surface roughness was 0.5nm.

Introduction

The development of IC technology follows Moore law, and the line width is thinner and thinner. At present, 65nm technology had been mass production, and 45 nm technology are researched successfully. The diameter of silicon substrate was developed from 200mm to 300mm. With the rapidly developing of ULSI fabrication technique to high integration and low cost, the key technology is how to ensure the higher flattening, lower roughness and lower cost[1-4].

Described generically as the process of smoothing and/or planning synergistically aided by combined chemical and mechanical effects, CMP has been widely and effectively applied to a variety of materials in last two decades, including semiconductors, crystals, glasses, special metals, plastics and computer disks. CMP is becoming popular because the advantages of CMP over other surface finishing techniques are great: little or no surface and subsurface damage, elimination of surface defects, high removal rates, high efficiency, and low cost. And it has been confirmed that CMP is the unique and best method for the global planarization of ULSI multilayer interconnection [4-7].

However, CMP is related to mechanics, hydrodynamics, physical chemistry, material tribology and other subjects, its kinetics process and control process is still not clear. The whole process for CMP is the chemical mechanical complex process, i.e. the relative velocity between the pad and the wafer surface, stable pressure and temperature, certain pad and slurry with abrasive and additives [8-10]. How to confirm the kinetics process and find the kinetic control process is the key to improving CMP surface machining quality of silicon substrate. In this paper, the CMP mechanism process of silicon wafer was analyzed. The influence of surfactant on silicon surface quality during CMP process was also discussed, and the contrast experiments were processed. At the end of silicon CMP, the water polishing was processed using de-ionized water with surfactant, and the optimal surface was gotten.

CMP Mechanism Analysis of Silicon Substrate

Chemical mechanical polishing kinetics process

Research indicates that CMP is a complicated multiphase reaction process, eg. ULSI silicon substrate polishing. Under the action of pressure, revolve and mechanical abrasive, reagent and reactant is more adjacent and easier to react. During the revolving, the energy of reactant particles increases, and the impact between particles also increases, then the particle proportion up to activation energy is more. During the CMP process, the main process is chemical reaction and diffusion in addition to the mechanical action of invariable pressure, velocity, pad and abrasive. The slower process controls removal rate. In order to ensure better polishing effect, the mass transport process must be accelerated, which can lead to high smooth, low damage, high rate and achieve global planarization. Mass transport includes two aspects. First the reactant access to wafer surface in time, and the active component in polishing slurry reacts with the atoms of wafer surface. The second is that reaction product separate from surface of wafer rapidly, and new surface is exposed to polishing slurry. So how to accelerate the mass transport is the key factor of CMP process.

CMP kinetic process was thoroughly investigated, and it was generalized the seven steps, as follows: (1) Reactant molecular diffuses from main flow polishing slurry to unfinished wafer exterior surface; (2) Reactant molecular diffuses from exterior surface to internal surface; (3)Reactant is adsorbed on unfinished wafer surface; (4)Chemical reaction occurs between reactant and unfinished material on surface, and the chemical reaction product is formed;(5)The chemical reaction product desorbs from wafer surface; (6)The chemical reaction product diffuses from internal surface to exterior surface;(7)The chemical reaction product diffuses from exterior surface to main flow polishing slurry.

At equilibrium state, the seven steps rate is equal. The removal rate is controlled by slower process which resistance is the largest. So decreasing the resistance of slower process, the total removal rate can be accelerated. If chemical reactive rate is smaller, the total removal rate of wafer is also small. Furthermore, the surface degree of finish is not good. On the contrary, even if chemical reaction is very rapid, but desorption is very slow, the total removal rate is not good. Because the chemical reaction products cannot separate from the surface of silicon immediately, the active component in polishing slurry cannot expose and react with the atoms on the new surface of wafer, which holds up chemical reaction. Improving the chemical reaction and the mass transport may enhance the removal rate; reduce damage layer and low roughness.

Mechanism of Chemical reaction

It is well known that SiO_2 plays a mechanical abrasive role during the silicon substrate CMP with alkali slurry. The reaction is[11]:

$$Si+2MOH+H_2O \rightarrow M_2SiO_3+2H_2 \uparrow \quad (1)$$

That does not accord with the experiment fact. The presence of the silica sols influences the removal rate greatly. The adding of 1/10000 silica sols can make the removal rate improve to 60- 80μm/h. Obviously, it is not only because silica sols improve the machinery effect of grind, but also the other influence.

According to our experiments，the reaction may exist as follow:

$$SiO_2 + Si \xrightarrow{pressure\ +OH^-} 2SiO \quad (2)$$

$$2SiO + O_2 \rightarrow 2SiO_2 \quad (3)$$

The product SiO of equation (2)is unstable, it is oxidized rapidly by the oxygen dissolved in slurry to produce stable SiO_2. SiO_2 grows to form colloidal particle with electric double layer on the surface which attracts part of OH^-, and the pH is decreased slightly. The reaction (2) and (3) can reflect the essence of high removal rate and not changing pH value during the CMP process.

In addition, the product M_2SiO_3 of the reaction (1) may decompose because of the strongly alkaline of MOH and weak acid of silicic acid, the reaction is:

$$M_2SiO_3 + 2H_2O \xrightarrow{pressure} H_2SiO_3 + 2MOH \quad (4)$$

This induces the slowly decrease of concentration of OH^- and illustrates that the organic alkali has the buffer action on pH modulating and catalytic action on the reaction. The total reaction may be:

$$Si + 3H_2O \xrightarrow{pressure} H_2SiO_3 + 2H_2\uparrow \quad (5)$$

Surface states after CMP

Surface states after CMP are related to the mass-transfer in the processing. Because of the low rate of the mass-transfer, the chemical reactive rates are not uniform. It can not efficiently realize the selectivity between top and low areas due to the non-uniformity of local rates, so it is difficult to achieve the global planarization with low roughness. To enhance the effect of the mass-transfer, we can take two aspects as follows: (1) Enhance abrasive concentration and make sure that dispersion is still low; (2) Add non-ionic surfactant which will not bring negative influence to CMP. The increasing of abrasive concentration and the adding of surfactant can enhance the transmit velocity between reacting substances and reaction products, and also increase the removal rate and the concave-convex selectivity.

Effect of Surfactant on Si CMP process

The complex non-ionic surfactant added to CMP slurry may have effects of improving suspension of the abrasive, selectivity of concave area and convex area, reducing the surface tension and accelerating mass-transfer.

Abrasive suspension

The surfactant can act as dispersant to improve the uniformity and stability of the slurry. According DLVO theory [12], the stabilization of sol lies on the attraction or static repulsive force between the particles. The non-ionic surfactant can form special adsorption layer on the surface of particle to decrease the attractive potential energy and a new type of repulsive potential energy is produced after the particles surrounded by the surfactant (shown in Fig.1)to improve the uniformity and stability to insure the grinding and transportation homogeneous to finish the agitation and carrying the reactant and reaction product. The abrasive silica sol for the experiment is measured by TEM, shown as Fig. 2, particle size is 40nm.

Ex. 1016, Page 577

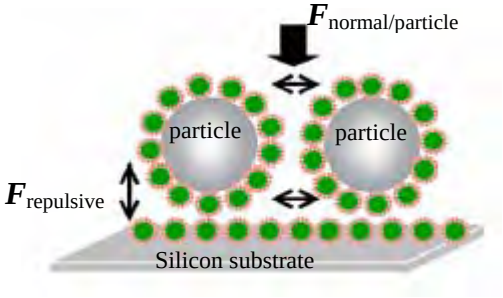


Fig.1 suspension effect of surfactant on particle

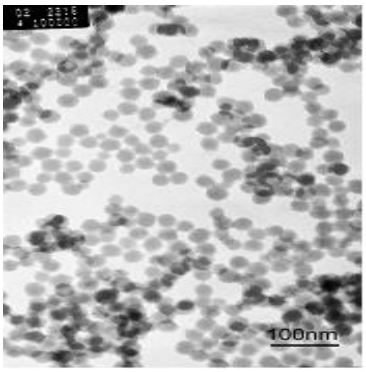


Fig.2 The pictures of Silica sol taken by TEM

Selectivity between high area and low area

The device features increasing smaller in process 0.15μm and beyond the well selectivity must be obtained. The selectivity of Si CMP is between high area and low area, to achieve the final planarization the removing rate of high area should be increased and the low area should be decreased. Surfactant can stick to silicon film easily; it also can separate abrasive and prevent the settle-down of slurry.

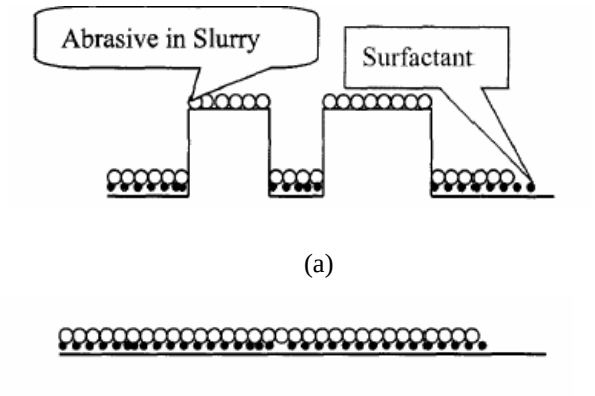


Fig. 3 CMP processing with surfactant improving selectivity

The complex surfactant can adsorb on the polished surface to separate the substrate surface and abrasive, see Fig.3 (a). But on the high area the abrasive can touch the wafer surface because the high pressure prevents the surfactant to stick on wafer. So the rate of low area is reduced and the rate of high area is increased. The surface will be planar quickly and then the surfactant stick to the wafer surface uniform, avoid the abrasive particles adsorb on the surface that is benefit to after CMP cleaning, see Fig.3 (b).

Roughness control

The surfactant molecules can be adsorbed physically on the solid surface because of its big molecule and can accelerate the mass transfer of the reaction and production to increase the planarization.

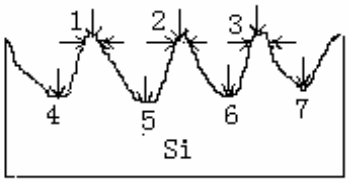


Fig. 4 Difference of the transportation rate

If reaction products are transmitted quickly from the surface while reagents are passed over to the surface in time, it will appear high speed difference and low roughness is realized.In Fig.4, area 1,2,3 are damage apices, they have removing rate around besides the rate adown. But at low area 4,5,6,7 there are only the rate adown. The high rate of apices leads to low damage and high planar surface.

Ex. 1016, Page 578

Particle removal

Particles are easily adsorbed on the polished surface and need to be removed by after CMP cleaning process; we adopted water polishing process after CMP with 1% FA/O surfactant and no pressure. Surfactant molecules are easier to be adsorbed on the polished surface, it can permeate into the adsorbate and develops in deep as a wedge in the interface (see Fig.5) and the particles are floated in the water to be removed and substituted by the surfactant (see Fig.6). In this case around the particles a layer is formed by adsorption of surfactant, in result a barrier is developed between the wafer and adsorbate. Thus the detrimental chemical adsorption which is difficult to clean can be transformed into the physical adsorption which is easy to be washed out.

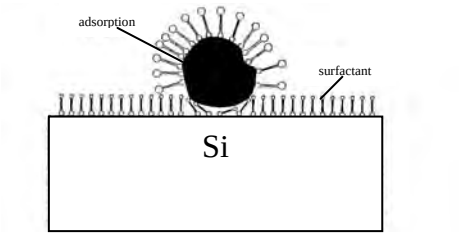


Fig. 5 Strong osmosis of the surfactant

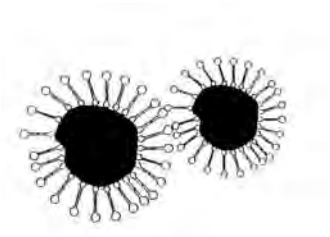
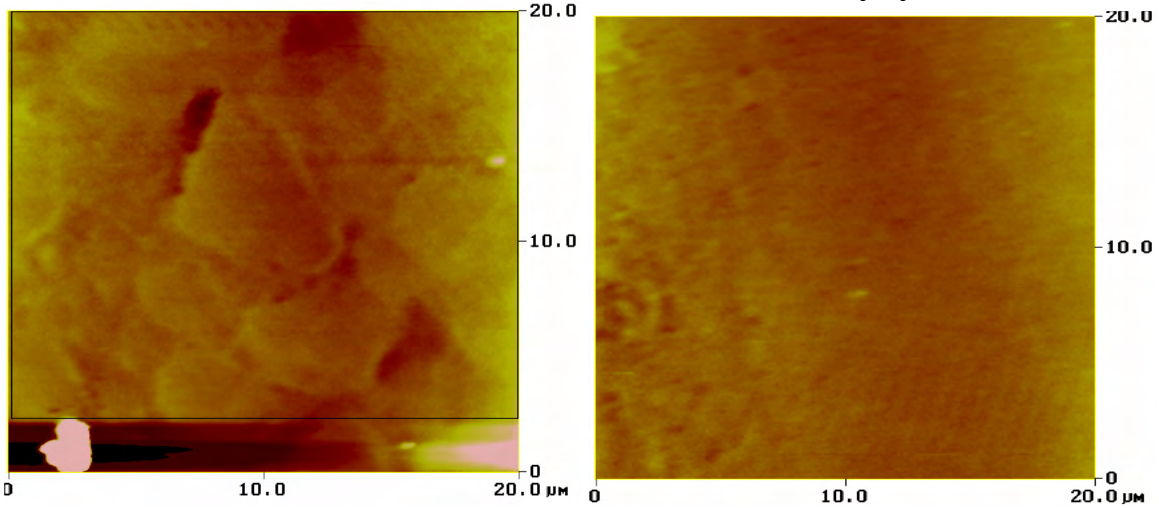


Fig. 6 The particle in the solution is surrounded by the surfactant

Experiment and result

The CMP experiments were accomplished on the polisher C6382I-W/YJ， with pressure 0.05Mpa, platen speed 60(RPM), polish pad IC8000, silicon wafer diameter Φ150mm,(100). Other parameters of the CMP processing are shown as follows: Abrasive silica sol concentration 5%, pH value 10.8 regulated by organic alkali, FA/O surfactant 1%, slurry flow: 200ml/min, temperature 25℃, following CMP process water polishing is taken with 0.5% surfactant in water and no pressure. The roughness (shown in Fig7.) was tested after CMP with and without surfactant in the CMP slurry by contrast.



(a)Ra= 0.9 nm without the surfactant (b) Ra= 0.5nm with the FA/O complex surfactant

Fig 7 The roughness of silicon surface after CMP

From the result, we can see that low roughness was realized by the surfactant in the slurry and Ra is 0.5nm, there is no particle on the surface.

Ex. 1016, Page 579

Conclusion

CMP kinetic process was generalized the seven steps, the removal rate and the surface degree of finish were controlled by the chemical reaction rate and the reactive product separated from the surface of silicon. The complex non-ionic surfactant added to alkali CMP slurry can improve suspension of the abrasive and selectivity of concave area and convex area, reduce the surface tension and accelerate mass-transfer. The CMP experiment showed that with alkali slurry, the FA/O surfactant can realize the low roughness to Ra 0.5nm on 20×20 μm.

Acknowledgments

The authors would like to thank the Specialized Research Fund for the Doctoral Program of Higher Education of China(20050080007) and Natural Science Foundation of China (No 10676008) for the financial support of this study.

References

1. Q.G.Zhou, Equipment For Electronic Products Manufacturing，10，1（2005）
2. J.J Vlassak, Journal of the Mechanics and Physics of Solids，52（4）, 847(2004)
3. Yong-Jin Seo, Sang-Yong Kim and Yeon-Ok Choi, Materials Letters，58（15）, 2091(2004)
4. Liu Yuling and ZhangKailiang, Microelectronic Engineering, 66(1-4),438(2003)
5. L. Charns and M. Sugiyama, Thin Solid Films, 485,188(2005)
6. Honglin Zhu, Luiz A. Tessaroto and Robert Sabia, Appl.Surf.Sci, 236,120(2004)
7. Xing Zhe, Liu Yu ling and Tan Baimei, Chinese Journal of Semiconductors, 25,1726(2004)
8. Zhang Kailiang, Liu Yu ling and Wang Fang, Chinese Journal of Semiconductors, 25,115(2004)
9. Guanghui Fu and Abhijit Chandra, Thin Solid Films, 474, 217(2005)
10. Yongwu Zhao and L. Chang, Wear, 2003,254 : 332
11. Chen Yang, Chen Zhigang and Li Xiazhang, Lubrication Engneering 176,120(2006)
12. Shen Zhong, Beijing: Chemical Industry Press, 2004.p8.

Ex. 1016, Page 580

ECS Transactions, 18 (1) 553-557 (2009)
10.1149/1.3096501 © The Electrochemical Society

Poly-CMP Slurry for Flash Memories

Judy Jing[a]; Bryan Chiu[b], Travis Huang[b], Arthur Hsu[b], Andy Yang[a]; Chris Yu[a]

[a] Anji Microelectronics (Shanghai) Co., Ltd., Shanghai, 201203, China
[b] Applied Materials Taiwan, Taiwan

Abstract

IC market demand for flash memories showed a significant growth trend in recent years. In manufacturing of flash memories, one of the most important processes is forming floating gate or control gate, which is greatly influenced by gate-poly CMP performance. To prepare a floating gate or control gate, a poly-silicon (poly) layer will be deposited on the active area (AA) , which is between the STI trench structures, typically HDP oxide. Then through CMP technique, the overburdened poly will be removed, with no poly residue on the top of HDP surface and leaving a certain thickness of AA poly whose top surface is planar with HDP top surface. However, in the case that a poly layer and a silicon dioxide layer are being polished, the removal rate of poly-silicon will tend to be much higher than that of silicon dioxide, resulting in recess between the STI trench structures and a non-planarized surface. To overcome this issue, we report here a novel poly CMP slurry that has the self-stopping capability and good planarization efficiency, with which the excellent surface planarity, poly dishing and poly residue clearance were obtained on 300 mm patterned wafers.

Introduction

The demand in IC market for flash memory is presenting a significant growth since several years ago, which speeds up the scaling down of flash memory compared with that of DRAM recently[1]. While some technologies of producing flash memory are not mature enough to enable an efficient and high performance manufacturing. In manufacturing flash memories, one of the most important processes is forming floating gate or control gate, whose final performance is greatly influenced by Gate-poly CMP performance[2]. As shown in Figure 1, to prepare a floating gate or control gate, the removal of an overburdened poly film layer deposited on HDP oxide film is needed, leaving no poly residue on the top of HDP surface and a certain thickness of Active Area (AA) poly whose top surface is ideally planar with HDP top surface. Besides, the variation of remaining poly film thickness between cell and periphery region is expected to be minimized. Unfortunately, these requirements in manufacturing flash memory can not be realized directly through those commercialized conventional CMP slurries such as oxide CMP slurry, because the planarization efficiency of commercialized conventional slurries are low and the removal rate of poly silicon as well as selectivity (poly/oxide) of commercialized conventional slurries are high. To our knowledge, most of current processes of gate-poly CMP for flash memory do not meet performance requirements. Therefore, a combination of CMP process with other processes such as etching post-CMP

Ex. 1016, Page 581

has to be used as a solution. In this paper, Anji poly CMP slurry with a self-stopping polishing capability was introduced to the application of flash memory gate-poly CMP. The polishing performances in terms of poly and HDP removal rates, self-stopping behavior and poly residue clearance as well as topography planarization were investigated.

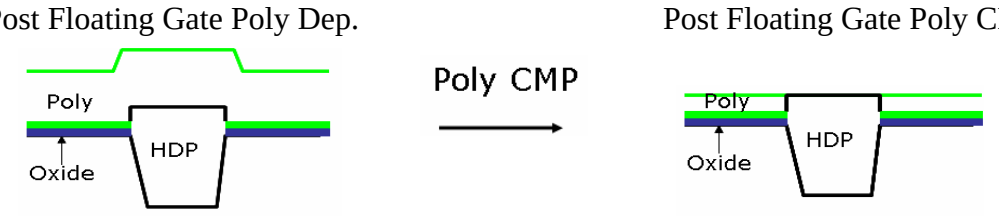


Figure 1. Scheme of gate-poly CMP for flash memory

Experimental

The polishing experiments were carried out in fab with the AMAT Reflexion LK oxide polisher and 300 mm blanket and patterned wafers. The polishing pad was IC1010 pad, and the pad conditioner was 3M A166 disk. In each polishing experiment, the poly slurry from Anji company was used with a slurry flow rate (SFR) of 200 ml/min. After polishing, the poly film thickness and poly residue status were measured with optical and scanning electron microscope methods, respectively.

Results and Discussions

1. The Poly and HDP Removal rate on blanket wafers

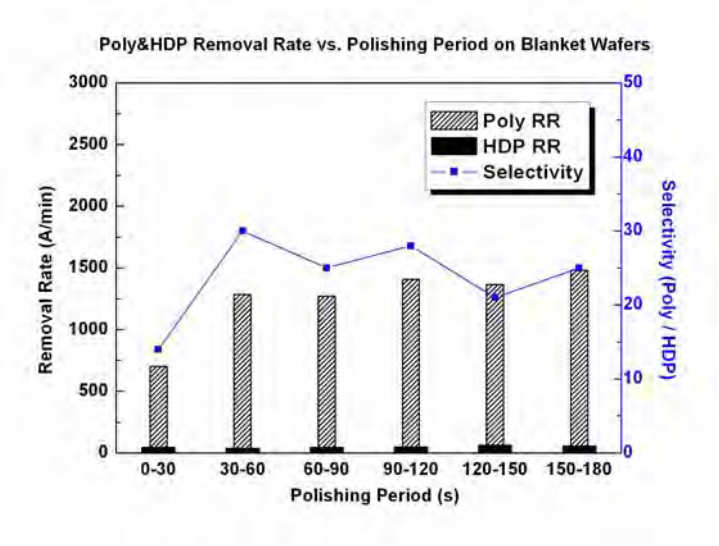
Figure 2(a) shows the poly and HDP removal rate on 300 mm blanket wafers at different polishing periods obtained with Poly slurry from Anji company at 1.5 psi main polishing down-force. In Figure 2(a), the Poly and HDP removal rate are stable at around 1360 A/min and 50 A/min, respectively, while the poly removal rate of first polishing period (0-30 s) is only about 700 A/min. The lower poly removal rate of 0-30s polishing period should be attributed to the passivation effect from thin native oxide film covered on poly film surface because the native oxide material is much more difficult to be removed than the poly material. Therefore the removal rate selectivity of poly to HDP is 14 for 0-30s polishing period and 20~30 for the stable polishing periods, respectively. Normally, the Poly removal rate of Gate poly CMP needn't high because the poly removal amount required in Gate poly CMP is lower than that in plug or capacitor poly CMP in DRAM manufacturing process**.** The low poly removal rate is not difficult to be obtained by slurry formulation tuning, however, the removal rate selectivity of poly to HDP must be selected carefully considering the balance between production throughput, poly residue clearance, poly dishing and polishing end point performances. These performances can be improved at the same time through a self-stopping polishing process with Anji poly slurry, which will be explained in details subsequently.

2. The self-stopping polishing on pattern wafers

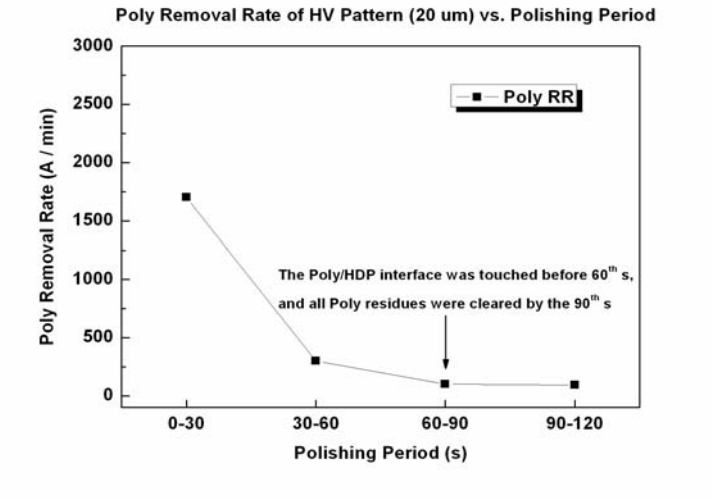
Figure 2(b) shows the poly removal rate of different polishing periods with Anji poly slurry based on the remaining poly thickness measurements of 20 um HV pattern in 300 mm patterned wafers. Actually, the curve plotted in Figure 2 (b) depicted a self-stopping polishing course on patterned wafer. That is, with the increase of the

Ex. 1016, Page 582

polishing time, the poly removal rate of 20 um HV pattern decrease obviously at the beginning of the polishing and reach a low but stable value finally. The self-stopping polishing enables both a high throughput in Gate poly CMP process due to the fast poly removal at the initial polishing stage and a good polishing stop as well as low poly dishing result due to the very low poly removal rate at the end polishing stage. At the same time, the HDP removal rate is high enough to ensure the capability of clearing poly residue in the non-planar HDP surface region, considering that the poly polishing can be stopped by itself rather than by HDP polishing. In Figure 2(b), only small amount of poly residue remains after 60 s polishing and it can be fully cleared after 90 s polishing. It is reasonable to speculate that the combination of self-stopping polishing with optimized process steps, parameters, and ISRM end point detection would result in a much improved polishing result.



(a)



(b)

Figure 2. Removal rate obtained on (a) blanket wafers and (b) patterned wafers

3. The capability of tuning removal rate profile

The capability of tuning removal rate through varying the CMP down-force or rotation speed is important to enable the realization of the within wafer non-uniformity

Ex. 1016, Page 583

(WIWNU). Therefore the investigations on sensitivity of removal rate to the process mechanical parameter—down-force were carried out. Investigation results are shown in Figure 3. According to Figure 3, we conclude that with Anji poly slurry, the poly removal rate profile on both blanket and patterned wafer can be easily tuned by changing mechanical parameters of the process, and the desirable WIWNU post CMP is definitely available through the optimization of CMP process.

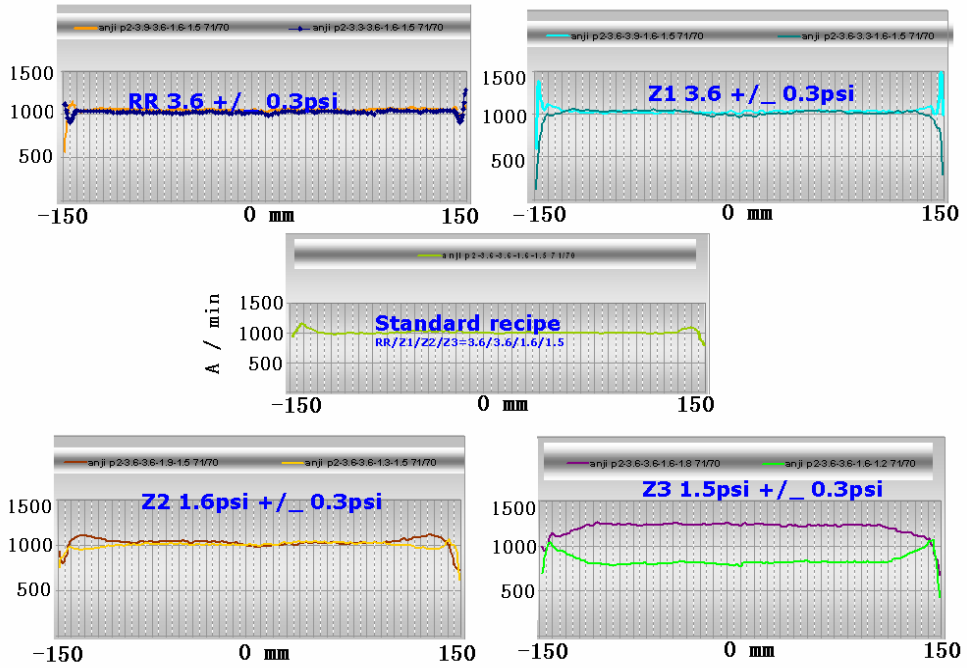


Figure 3. The removal rate profiles obtained with standard recipe, and recipes only changed retaining ring (RR) or zone 1 (Z1) or zone 2 (Z2) or zone 3 (Z3) down force compared with standard recipe. Standard recipe down force (psi): RR/Z1/Z2/Z3 = 3.6/3.6/1.6/1.5.

4. The capability of planarizing the surface topography

As mentioned in introduction part, the planarization efficiency is expected to be high for Gate-poly CMP to minimize the variation of poly remaining thickness between cell and periphery region and the poly dishing. Figure 4 shows the capability of planarizing the surface topography in floating gate structure of 300 mm wafer with Anji poly slurry. In terms of Figure 4, the ranges between the poly remaining thicknesses of cell and 20 um HV region are decreased obviously with the increased polishing time (PT). The poly dishing of 20 um HV also has the same trend. Thus after a 30 s over-polish (60 s to 90 s) for poly residue clearance, the topography variation including range and dishing were minimized to a rather low level.

Evaluation item	Residue margin(Iso.)	Remain poly THK C.S nm			Dishing
		Cell	20um(HV)	Range	20um(HV)
Anji SFR=200ml/min PT 30" (Array)	All NG	85 (83~87)	51 (41~61)	34 (max:46)	13.5
Anji SFR=200ml/min PT 60" (Array)	Cell slightly NG	62 (60~64)	37 (30~44)	25 (max:34)	13
Anji SFR=200ml/min PT 90" (Array)	All clear	56 (54~57)	33 (30~35)	23 (max:27)	2.5

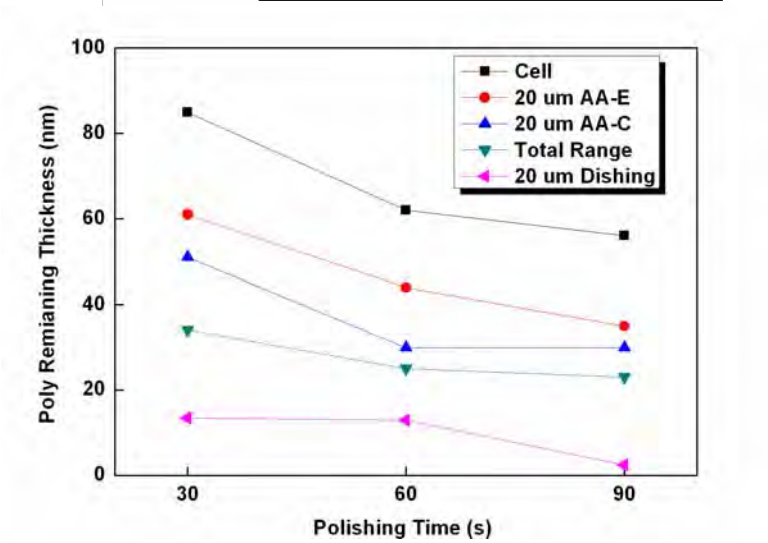


Figure 4. The capability of planarizing the surface topography with Anji poly slurry. (NG = Not Good, THK = Thickness).

Conclusions

With Anji poly slurry, the self-stopping Gate-poly CMP process was performed successfully on 300 mm patterned wafers. The self-stopping polishing using Anji poly slurry resulted in excellent performance at multiple aspects such as the polishing throughput, polishing non-uniformity and poly residue clearance as well as surface planarity post CMP polishing. It is believed that improved performances should be obtained through further optimized CMP process.

Acknowledgments

The authors would like to thank all members of Anji microelectronics Co., Ltd. for their kind support.

References

1. International Technology Roadmap for Semiconductors, 2007
2. Slotboom, M., Goarin, P., Akil, N., etc. ate isolation technology for compact poly-CMP embedded flash memories European Solid-State Device Research, 2003. ESSDERC apos, 03. 33rd Conference, 16-18 Sept. 2003 Page(s): 159-162.

Ex. 1016, Page 585

ECS Transactions, 18 (1) 559-564 (2009)
10.1149/1.3096502 © The Electrochemical Society

Synthesis of Spherical-like Ceria Particle with PVP as Assistant Agent and Its CMP Performance on Shallow Trench Isolation

Zefang Zhang [a,b,c], Weili Liu [a,c], Zhitang Song [a,c], Xiaokai Hu [a,c]

[a] State Key Laboratory of Functional Materials for Informatics, Laboratory of Nanotechnology, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China
[b] Graduate School of the Chinese Academy of Sciences, Beijing 100049, China
[c] Shanghai Xinanna Electronic Science & Technology Co., Ltd., Shanghai 201506, China

Abstract

Chemical Mechanical Polishing (CMP) has become a widely accepted global panarization technology. Abrasive is one of the key elements during CMP process. Du to its high selectivity, ceria particle is often used as abrasive in shallow trench isolation (STI) CMP slurries. But its irregular shape and agglomeration is not acceptable. In the present work, the spherical-like ceria nano-particles were successfully synthesized by homogeneous precipitation of ammonium cerium nitrate and urea, with Polyvinylpyrrolidone (PVP) as assistant agent. The precipitated ceria particles were characterized by means of XRD, FTIR, SEM and EDS as well as Zeta potential test. The results indicated that these nano-particles had a good dispersibility in water media without addition of any extra dispersing agent. Then, the CMP performance of these nano-particles on STI was investigated. The results indicated that they exhibited high removal selectivity (10.8:1) between silicon dioxide and silicon nitride without any additives.

Introduction

Ceria as an abrasive has been used to glass CMP for centuries. Recently, ceria has become very important in STI CMP, due to the good removal selectivity and the high polishing efficiency of the oxide film. Several works had synthesized of ceria and investigated its CMP performance. Ming-Shyong Tsai (1) synthesized nano grade cerium oxide via homogenous precipitation of ammonium cerium nitrate and urea and investigated its CMP performance on STI. The best polishing performance that the selectivity between silicon dioxide and silicon nitride is 10.2 have been found. Taekyung Yu et al. (2) prepared water dispersible ceria nanocrystals by a simple sol-gel reaction of cerium salt in the presence of oleylamine. The aqueous dispersion of the product was successfully used as CMP slurry and it exhibited high removal selectivity between silicon oxide and silicon nitride at pH 7. Ye-Hwan Kim et al. (3) synthesized nano-sized ceria particles via a solid-state displacement reaction method and controlled their crystalline structure by regulating the oxygen partial pressure at the reaction site on the precursor. In Ye-Hwan Kim's study, the effect of the crystalline structures of nano-sized ceria particles on STI CMP was investigated. However, the problem of the dispersibility and morphology of ceria abrasive have not been resolved. In this study, the spherical-like

Ex. 1016, Page 587

ceria nano-particles with good dispersibility were prepared by modifying the process described in reference (1) through adding the PVP during the reaction process. The difference between the present work and the literature reports is PVP using as assistant was added to control the morphology of the ceria nono-particles and enhance its dispersibility and stability. And these ceria nano-particles were successfully used as an abrasive for STI CMP, the slurry exhibited high removal selectivity between silicon oxide and silicon nitride.

Experimental

Synthesis and Characterization of Ceria Nano-particles

All the reagents were of analytical grade purity and were used without further purification. 1.5 g PVP was dissolved in 150 mL Deionized (DI) water and stirred magnetically to get transparent solution, then 8 g ammonium cerium nitrate and 3.5 g urea was added and stirred magnetically getting red mixed solution. The red mixed solution was heated and refluxed at 100℃ for 10 h. The precipitate was cooled and washed three times with DI water and finally anhydrous alcohol for one time. The final product was dried at 80 ℃ for 3h. The structures of these ceria particles were characterized by X-ray diffraction (XRD, D/max-2200PS, Rigaku) and Fourier transform infrared (FTIR). The morphology of the products was examined by field emission scanning electron microscope (SEM, S-4700, Hitachi). The stability of ceria particles was studied through a Zeta potential analyzer (Nicomp 380/ZetaPALS, PSS).

Chemical Mechanical Polishing Process

The polishing slurry containing 1 wt.% ceria was prepared, and the pH slurry was fixed to 7. Silicon wafer with 1000 nm SiO_2 film and silicon wafer with 300 nm Si_3N_4 film were polished using a CMP tester (CETR, CP-4) with an IC 1000/Suba IV stacked pad (Rodel). The polishing process parameters were set as follows: pad rotation speed 100 rpm, wafer rotation speed 100 rpm, down force 3 psi, feed rate of the slurry 60 mL/min and polishing time 1 min. The difference between the thickness of each film before and after polishing was measured by SEM and defined as the material removal rate (MRR).

Results and Discussions

Structure and Morphology of As-synthesized Ceria Nano-abrasive

Fig. 1 shows the XRD pattern of the as-synthesized ceria nanoparticles. All the diffraction peaks correspond to the (111), (200), (220), (311), (222), (400), (331), (420), and (422) planes, which can be indexed to the face-centered cubic structure for CeO2 (space group Fm3m (225)) with a lattice constant of $\alpha = 0.5411$ nm according to JCPDS No. 65-2975. No obvious peaks corresponding to cerium nitrate or other cerium oxides were observed in the powder pattern.

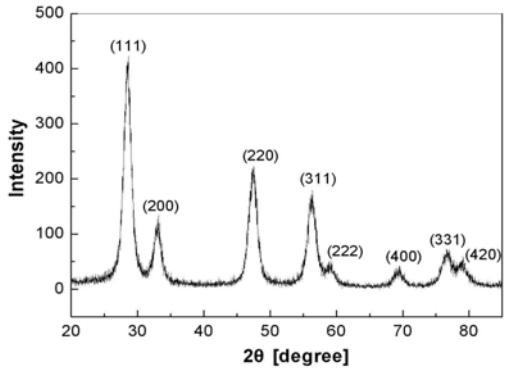


Fig. 1 XRD pattern of as-synthesized ceria nano-particles

More characteristics of as-synthesized ceria nano-abrasive are also observed in its FTIR spectra (Fig. 2). The band centered at 3436.53 cm^{-1} is assigned to O-H stretching vibration in -OH group. In addition, a series of absorption peaks at 2974.41 cm^{-1}, 1627.63 cm^{-1}, 1292.07 and 1114.65 cm^{-1} which are ascribe to $–CH_2$, C=O, C-N, respectively, show that ceria was in situ surface modified by PVP during the synthesis process.

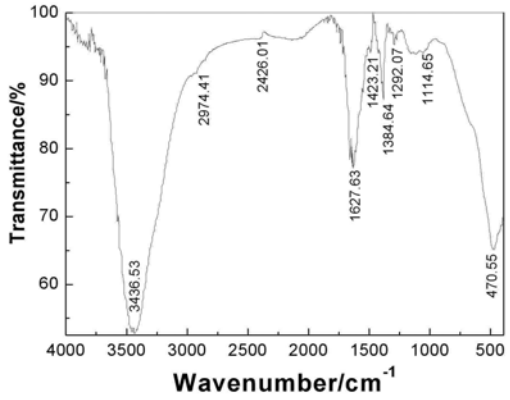


Fig. 2 FTIR spectrum of as-synthesized ceria nano-particles

The morphology of the products was investigated by the SEM techniques (Fig. 3). Images in Fig. 3 show that the sample was consisted of sphere-like particles with diameter of about 50 nm. It has been found that PVP play an important role in controlling the morphology of nano-particles (4). The components of the area in Fig. 3 analyzed by energy dispersive spectrum (EDS) are shown in Fig. 4, which indicate that the elements in the SEM photograph are Ce, O, N, C and Si. The Ce and O peaks may be attributed to as prepared ceria nano-abrasive, N and C peaks can be attributed to PVP on the surface of ceria nano-abrasive which have been confirmed by FTIR in Fig. 2, Si signal is from specimen support.

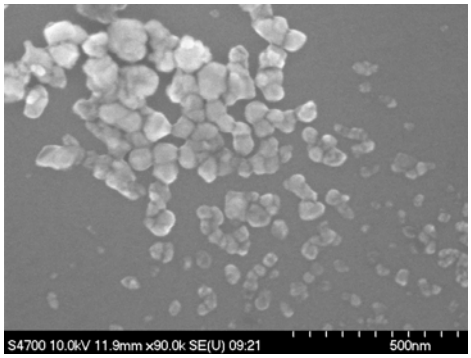


Fig. 3 SEM image of as synthesized ceria nano-particles

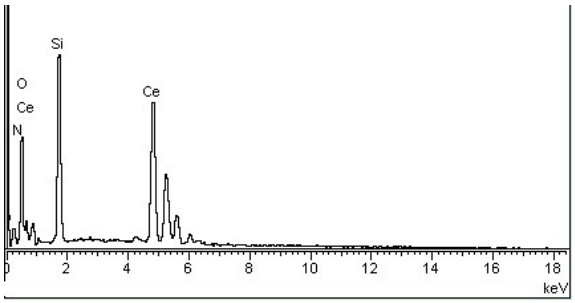


Fig. 4 EDS spectrum of as synthesized ceria nano-particles

Stability of As-synthesized Ceria Nano-abrasive

Fig. 5 shows the photograph of the slurry that was kept for more than two weeks. No precipitate was produced, which suggested that the dispersibility of the slurry was good. In order to further analyze the stability of the ceria slurry, the Zeta potential of the slurry was investigated and the result of which was seen in Fig. 6. As we know, the unstable region of colloidal abrasive is commonly considered to be from -30 mV to 30 mV. The average Zeta potential for as-synthesized ceria abrasive is 40.76 mV. It is more than 30 mV, which illustrated that the ceria are very stable. As can be seen form Fig. 2 that ceria nano-particles showed a strong and broad band at 3436.53 cm^{-1}, which is the result of Ce-OH that was located on the surface of the ceria nano-particles. These surface Ce-OH groups seem to allow these nano-particles to be water dispersible (3). Furthermore, the PVP on the surface of ceria can obstruct agglomeration of particles in slurry (5).



Fig. 5 Photographs of as prepared slurry that was kept for more than two weeks

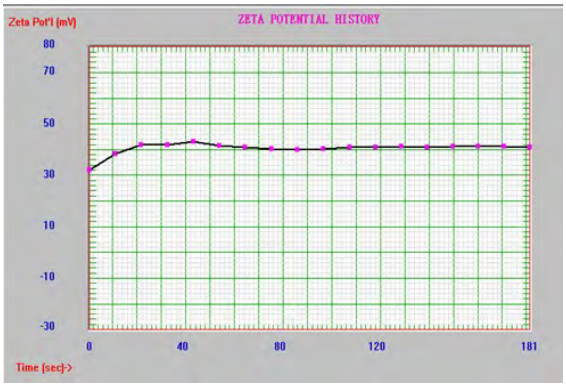


Fig. 6 Zeta potential of as-synthesized ceria nano-particles

Polishing Properties of As-synthesized Ceria Nano-abrasive

The CMP performance of the as-synthesized ceria nano-abrasive on STI was investigated, and the typical SEM cross section images of the SiO_2 wafer before and after polishing are shown in Fig. 6. The difference of the thickness of SiO_2 film in Fig. 6 between before and after polishing is the MRR of SiO_2 film, and the MRR of Si_3N_4 was obtained as the same. It can be seen from Fig. 7 that as-synthesized ceria nano-abrasive exhibited the MRR 238 nm/min and 22 nm/min for SiO_2 film and Si_3N_4 film, respectively. So the polishing selectivity between SiO_2 film and Si_3N_4 film is 10.8 without any additives. In general, the selectivity of commercial ceria was lower than those ceria (1). In addition, effect of different synthesis condition on its CMP performance will be studied in our further work.

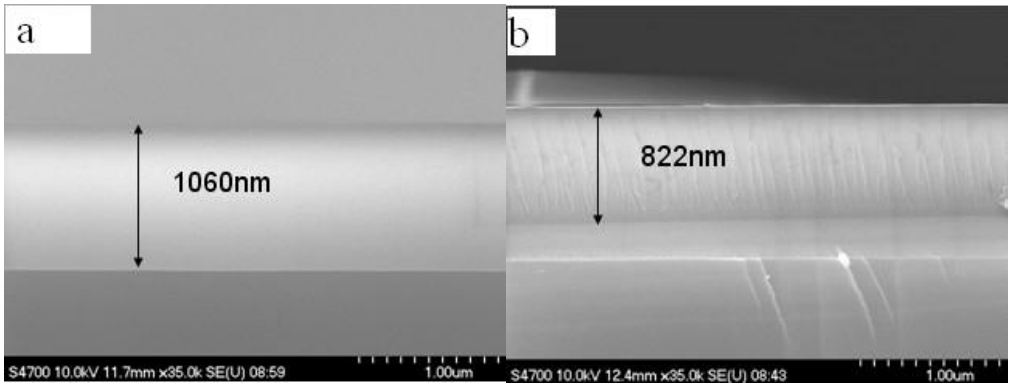


Fig. 6 Typical SEM cross section images of the SiO2 film before (a) and after (b) polishing

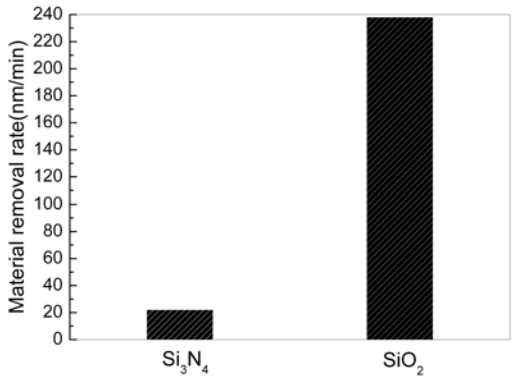


Fig. 7 MRR of different film

Conclusions

The spherical-like ceria nano-particles were successfully synthesized by homogeneous precipitation of ammonium cerium nitrate and urea, with PVP as assistant agent. The overall synthetic process is very simple, so an approach for large scale synthesis of ceria nano-particles was developed. Furthermore, the ceria nano-abrasive has a good dispersibility and can be kept for two weeks without any surfactant. The CMP test of these nano-particles on STI found that they exhibited high removal selectivity (10.8:1) between silicon dioxide and silicon nitride without any additives.

Acknowledgments

The work is supported by National Basic Research Program of China (2007CB935400 and 2006CB302700), National High Technology Development Program of China (2008AA031402, 2006AA03Z360), Science and Technology Council of Shanghai (0652nm003, 0752nm013, 07QA14065, 07SA08, 08111100301, 08111100303), and National Nature Science Foundation of China (60776058).

References

1. Tsai, M. S., *Mat. Sci. Eng. B-Solid*, **110**, 132 (2004).
2. Yu, T., Y. Il Park, et al., *Eur. J. Inorg. Chem.*, 855 (2008).
3. Kim, Y. H., S. K. Kim, et al, *Ultramicroscopy*, **108**, 1292 (2008).
4. Zhou, F., X. M. Ni, et al., *J. Colloid Interf. Sci.*, **307**, 135 (2007).
5. Sun, Y. G. and Y. N. Xia, *Science*, **298**, 2176 (2002).

CHAPTER 6

THIN FILM, ETCH AND PLATING

ECS Transactions, 18 (1) 567-574 (2009)
10.1149/1.3096503 © The Electrochemical Society

Advanced Process Control of $Hf_xSi_{1-x}O_2$ Composition using Co-injection Atomic Layer Deposition

L. D. Bartholomew[a], A. Bavin[a], T. Lazerand[a], V. Rao[a], M. Weldon[b] and L. Wang[b]

[a] Aviza Technology, 440 Kings Village Rd., Scotts Valley, California 95066, USA
[b] MetroSol, Inc., 2101 Donley Drive #101, Austin, Texas 78758, USA

A novel Atomic Layer Deposition (ALD) process has been developed in which chemical precursors are co-injected and simultaneously reacted. The process demonstrates hafnium silicate films of uniform composition that can be easily tuned from <30 to >80 % Hf concentration to meet specific device requirements with excellent thickness control proportional to the number of process cycles. The film has been fully integrated into a CMOS structure and provides repeatable parametric results over a wide range of film compositions. In addition we review a VUV metrology technique correlated to X-ray photoelectron spectroscopy that enables measurement of the hafnium vs. silicon composition in very thin films <50 Å, with sensitivity to lower hafnium concentration in the interfacial layer for films deposited directly upon silicon substrates.

Introduction

A key process challenge for sub-45 nm complementary metal oxide semiconductor (CMOS) devices is the adoption of high dielectric constant (κ) materials as the gate dielectric. At such small geometries, typical thermal silicon oxide (SiO_2) or silicon oxy-nitride (SiON) dielectrics are unable to meet the leakage requirements of the device without limiting the operating speed (1). By increasing the dielectric constant of the dielectric material it is possible to reduce the gate leakage while continuing to shrink the effective oxide thickness (EOT). Hafnium silicate ($Hf_xSi_{1-x}O_2$) is one of the leading gate dielectric materials for advanced CMOS high-κ schemes due to its high dielectric constant, band alignment to silicon, thermal stability and ease of integration into the manufacturing process (2). In addition, the hafnium composition of the film can be varied to change the physical and electrical properties of the material making it possible to tune the device performance to meet specific requirements.

Several techniques have been used to deposit $Hf_xSi_{1-x}O_2$ gate dielectrics, though Atomic Layer Deposition (ALD) most closely meets the application requirements. The ALD process uses discrete precursor and reactant pulses separated by a nonreactive gas purge to deposit a single atomic layer per process cycle. As the reaction is self limiting, ALD enables excellent control of film thickness and thickness uniformity: a key requirement for high-k gate dielectrics where film thicknesses are typically 15 to 20 Å. The technique also delivers high purity, high density films needed to achieve the required electrical performance. In a manufacturing environment, the process repeatability offered by ALD enables high process yields and reduces production costs.

ALD of binary component films such as $Hf_xSi_{1-x}O_2$ are typically deposited using the "nano-laminate" technique: sequential depositions of hafnium oxide and silicon oxide followed by a post deposition anneal to form a $Hf_xSi_{1-x}O_2$ film. This technique allows the

Ex. 1016, Page 595

use of a wide selection of precursors, but has a low deposition rate and requires an additional process step. Through careful selection of compatible precursors it is also possible to deposit $Hf_xSi_{1-x}O_2$ using a "co-injection" technique; hafnium and silicon precursors are pulsed at the same time followed by single oxidation pulse (Figure 1). Co-injection halves the number of process cycles and produces dense $Hf_xSi_{1-x}O_2$ films that are resistant to crystallization.

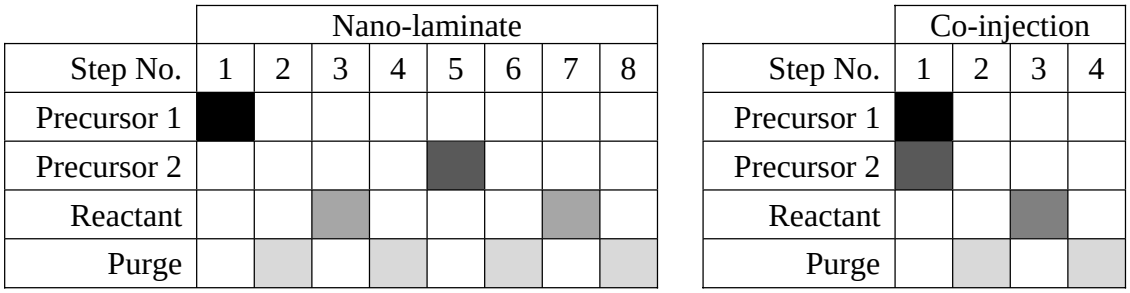


Figure 1. A comparison of nano-laminate and co-injection techniques illustrating the 50 % reduction in the number of cycles offered by co-injection ALD.

In this paper we demonstrate the capability of the co-injection ALD technique to deposit $Hf_xSi_{1-x}O_2$ films of uniform composition with Hf concentrations ranging from <30 to >80 %. In addition we review a Vacuum Ultraviolet (VUV) metrology technique correlated to X-ray photoelectron spectroscopy that enables measurement of the Hf vs. Si composition in very thin films <50 Å, with sensitivity to lower hafnium concentration in the interfacial layer for films deposited directly upon silicon substrates.

Experimental

The co-injection process was developed on a Celsior *f*xP™ ALD tool from Aviza Technology. The system employs a showerhead and radial pumping gallery to ensure uniform precursor and reactant delivery over a 300 mm Si wafer. A schematic diagram of the chamber and chemical delivery configuration is shown in Figure 2. The reaction volume is minimized to ensure rapid pulse and purge and is heated to 150 °C. The $Hf_xSi_{1-x}O_2$ was deposited using co-injected liquid Tetrakis-Di-Methyl-Amino-Hafnium (TDMAHf) and Tris-Di-Methyl-Amino-Silane (3DMASi) precursors of >99.999 % purity. The precursors were chosen for their low delivery temperature, chlorine-free chemistry and compatibility at process temperatures around 400 °C. A commercially available ozone delivery system was used to generate the ozone reactant with a concentration of 240 g/m^3 in oxygen. Argon (Ar) was used for the precursor carrier gas and purge gas. $Hf_xSi_{1-x}O_2$ films were deposited on reclaimed 300 mm Si (100) wafers measured to have 12 Å native oxide.

Several process conditions were evaluated to vary Hf content of the $Hf_xSi_{1-x}O_2$ film including process temperature, precursor temperature, process pressure and the flow rate of the Ar carrier gas.

Ex. 1016, Page 596

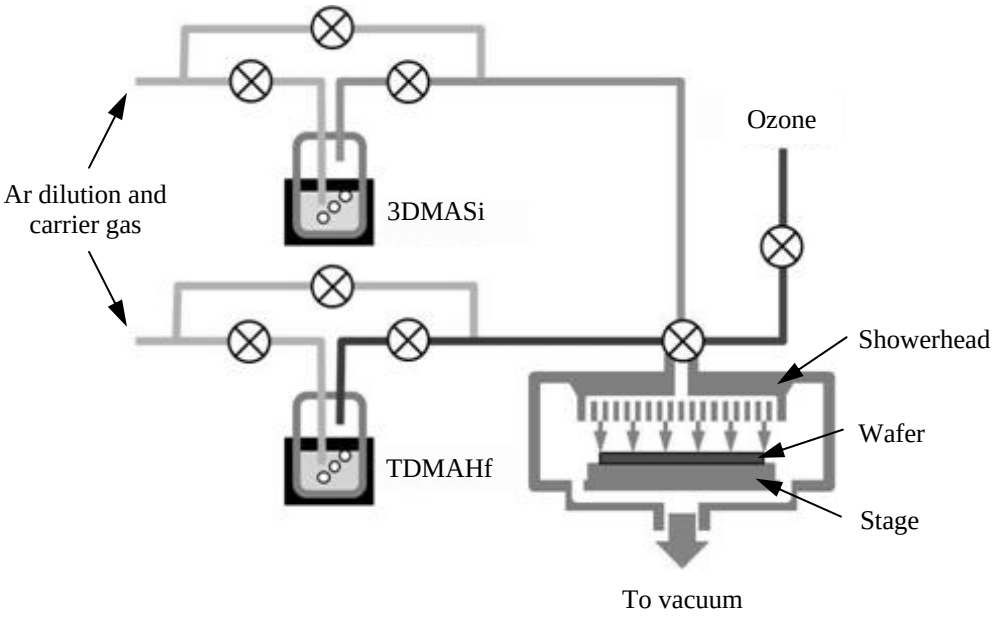


Figure 2. A schematic diagram of Celsior *f*xP™ ALD process chamber and chemical delivery system.

Characterization

The film thickness and Hf vs. Si concentration were measured at 49 points with an edge exclusion of 3 mm on a Metrosol VUV4000 system. This metrology technique enables accurate measurement on <50 Å films as proven through correlation with composition measurements performed by X-ray photoelectron spectroscopy (XPS) and TEM (3). In addition to the inherent sensitivity enhancement to thickness changes associated with the use of shorter wavelengths, the unique absorptive characteristics of materials in the Vacuum Ultraviolet (VUV) spectral regime enable simultaneous determination of thickness and composition. These unique optical properties are illustrated in the n and k plots of HfO_2 and SiO_2 shown in Figure 3.

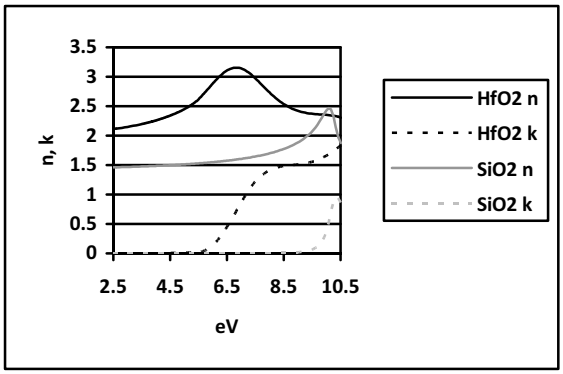


Figure 3. Plot of the optical properties (n & k) of HfO_2 and SiO_2, demonstrating the distinctive absorption properties found for many materials in the VUV.

The sensitivity of the VUV technique to lower hafnium concentrations and thickness in the interfacial layer ensures accurate measurements for films deposited directly upon silicon substrates. A raw reflectance spectrum representative of these samples and that of a Si substrate with a native oxide thin film are shown in Figure 4 to illustrate the

Ex. 1016, Page 597

sensitivity of the VUV spectral regime. Two model simulations illustrating small changes in oxide thickness are also superimposed to illustrate the inherent sensitivity to change found in the VUV.

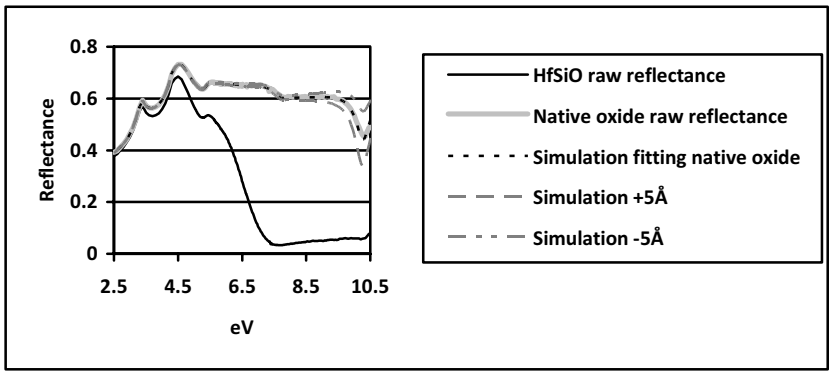


Figure 4. Representative 70 Å $Hf_{0.55}Si_{0.45}O_2$ wafer raw reflectance spectra, including Si Native Oxide raw reflectance data and three oxide reflectance simulations - one fitting the native oxide data as well as ones 5 Å thinner and thicker.

Optical thickness and refractive index were measured at 49 points with an edge exclusion of 3 mm on a KLA-Tencor F5x. Measurements of leakage current density and CET were performed on the as-deposited $Hf_xSi_{x-1}O_2$ films at 5 points with a contact area of 7.15×10^{-5} cm^2 using a Four Dimensions 3092-A Mercury Probe.

Results and Discussion

The stage temperature influences the incorporation of Hf and Si according to the relative reactivity of the precursors selected. The choice of 3DMASi allows greater incorporation of Si at lower temperature than alternative 4DMASi or TEMASi precursors due to their lower vapor pressure. TDMAHf has much higher vapor pressure than TEMAHf, allowing shorter pulse times for equivalent dosage without incurring as much decomposition from time at temperature on the wafer surface as both react at lower temperature than the compatible Si sources. Lowering the process temperature from 400 °C toward 300 °C, where TDMAHf has a more uniform self-limiting reaction, would decrease the Si content but sacrifice the electrical performance of the $Hf_xSi_{x-1}O_2$ film. The temperature of the liquid precursor can also be used to control the relative vapor pressures of the Hf and Si sources and thus the ratio of co-injected species and resulting composition of the film. However, while this technique may be suitable to tune the Hf content for a dedicated composition, adjustment of the precursor temperature may not be suitable for frequent composition changes due to the temperature stabilization time. Another simple recipe parameter that has an influence on Hf concentration is process pressure. As the 3DMASi is less reactive than TDMAHf and requires higher pressure to occupy surface sites for reaction, increasing the pressure reduces the Hf concentration.

The flow of the precursor carrier gas has a strong influence on the Hf content. Increasing the Ar gas flow through the respective bubblers carries a greater quantity of pure chemical vapor into the process chamber, and in return the ratio of the two precursors can be tuned by increasing one and decreasing the other. In an ALD system using rapid short pulses of chemical injection, the chemical vapor delivery is controlled to a greater extent by the relative delivery line pressure downstream of the bubblers as the flow path through the liquid is interrupted. The pulse time can also be used to adjust the

Ex. 1016, Page 598

composition in how it affects the delivery line pressures and the competitive saturation of the wafer surface between the two precursors.

By combining these various elements in the co-injection process recipe, the relative composition of Hf versus Si in the $Hf_xSi_{x-1}O_2$ film can be controlled throughout a continuous scale using the co-injection ALD process, whereas a nano-laminate process is limited to specific ratios according to the relative deposition rates and numbers of sequential individual HfO_2 and SiO_2 layers. Comparison of the Hf content data from the VUV technique with XPS shows good correlation as depicted in Figure 5.

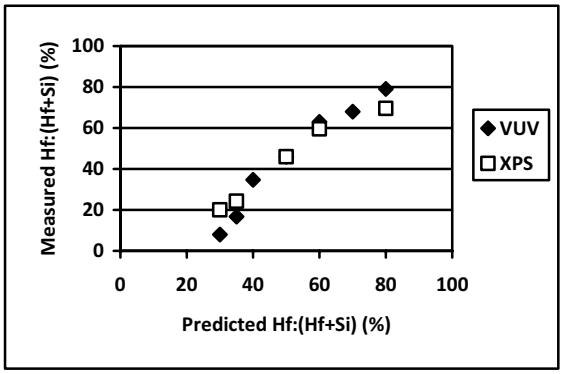


Figure 5. $Hf_xSi_{1-x}O_2$ films with linear control of Hf content from 30 to over 80 % obtained using a co-injection ALD process.

Since $Hf_xSi_{1-x}O_2$ films are deposited directly on Si substrates, the possible existence of an interfacial SiO_2 layer can be challenging for the analysis of the true film composition. As the deposited film thickness is reduced and the percentage of the interfacial oxide thickness relative to the total film increases, the measured VUV concentration of Hf decreases as expected (Figure 6).

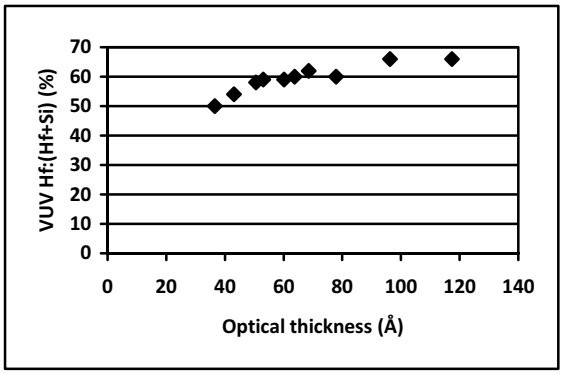


Figure 6. VUV %Hf concentration for $Hf_xSi_{1-x}O_2$ films of varying thickness deposited using the same ALD process recipe condition.

The bulk film Hf concentration at greater thicknesses where the interfacial oxide impact is minimal correlates to the XPS measured value which is typically analyzed at a minimum of 70 Å optical thickness. Therefore the VUV technique can distinctly analyze differences in composition at physical thicknesses below 50 Å as needed for actual device application.

Films with Hf compositions ranging from 35 to 80 % were analyzed to evaluate the potential impact on the physical and electrical characteristics of the $Hf_xSi_{1-x}O_2$. Linearity

tests resulted in a typical deposition rate of 0.75 Å/cycle with linear correlation between the optical film thickness and the number of process cycles for all compositions as shown in Figure 7. The linear nature of the plot is indicative that a true ALD process is being maintained for all Hf concentrations.

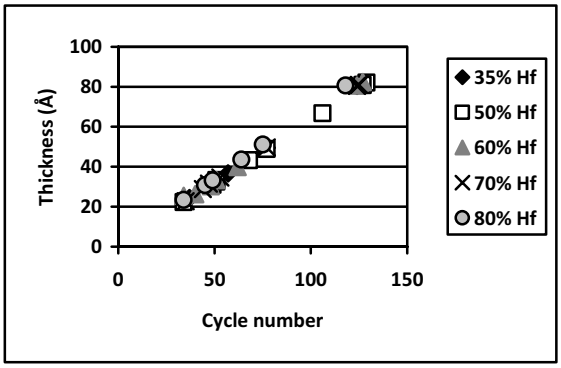


Figure 7. $Hf_xSi_{1-x}O_2$ process linearity at Hf concentrations from 35 to 80 %. All films demonstrate linear deposition properties indicative of a true ALD process.

As expected, an increase of the refractive index was seen with increasing Hf content. There was also a thickness dependence which correlates to the dilution effect of the interfacial oxide layer (Figure 8).

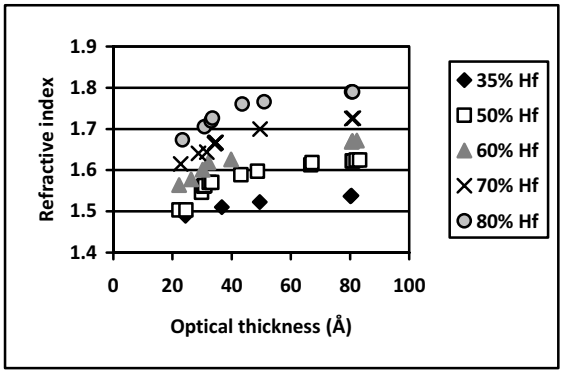


Figure 8. Refractive index of $Hf_xSi_{1-x}O_2$ films with 35 to 80 % Hf content. The reduction in RI for thinner films indicates dilution of the film from the interfacial oxide.

Higher Hf concentration films can be used at greater physical thickness to reduce leakage for the same capacitance equivalent thickness (CET). Figure 9 shows the leakage current for similar physical thickness films of varying Hf composition; for the same CET value, the leakage of higher Hf concentration films will be reduced even further as the physical thickness is increased.

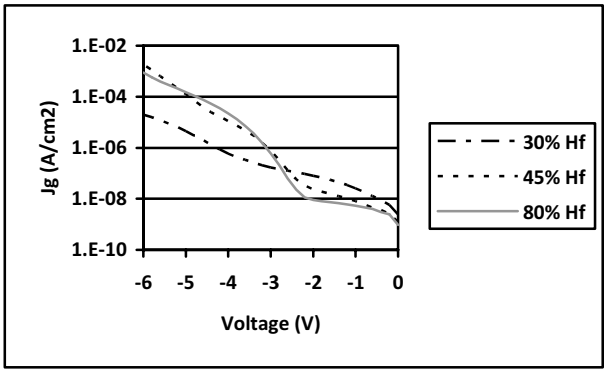


Figure 9. Leakage current density for different composition as-deposited $Hf_xSi_{1-x}O_2$ films of similar physical thickness.

For $Hf_xSi_{1-x}O_2$ films deposited on Si substrates having an interfacial oxide layer, the κ value of both the deposited high-κ film of thickness T_{phy} and the resulting interfacial layer of thickness T_{int} can be calculated using the following equations:

$$\kappa_{hi}*(EOT-T_{int}) + \kappa_{int}*Tint = 3.84*(T_{phy} + T_{int}) \quad [1]$$

$$(EOT-T_{int}) = (3.84/\kappa_{hi})*T_{phy} - (\kappa_{int}-3.84)*T_{int}/\kappa_{hi} \quad [2]$$

The thickness of the interfacial layer T_{int} is estimated to be the y-intercept from the optical thickness linearity plot. Plotting the measured CET values minus the thickness of the interfacial layer versus the physical thickness (optical thickness measured by the F5x minus the thickness of the interfacial layer) yields a slope and intercept. The resulting κ values for the dual layers of $Hf_xSi_{1-x}O_2$ and interfacial oxide are calculated to be:

$$\kappa_{hi} = 3.84/slope \quad [3]$$

$$\kappa_{int} = 3.84-(\kappa_{hi}/T_{int})*intercept \quad [4]$$

This calculation was used to estimate the κ_{hi} value of various composition $Hf_xSi_{1-x}O_2$ films. The results are shown in Figure 10.

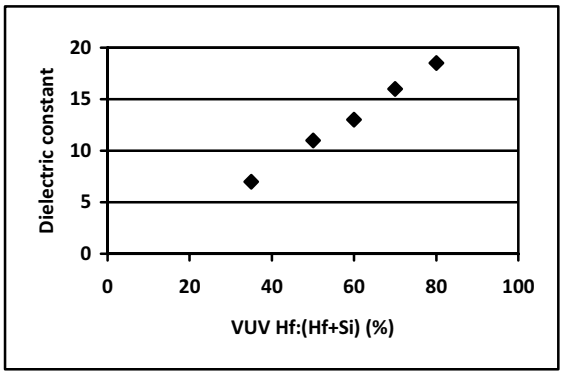


Figure 10. Calculated κ values for varying composition $Hf_xSi_{1-x}O_2$ films.

In advanced CMOS this change in dielectric constant can be used to tune the electrical performance of the gate. Figure 11 shows the gate leakage density and CET performance from a device incorporating co-injected ALD $Hf_xSi_{1-x}O_2$ films as a gate dielectric. It can clearly be seen that the increased Hf content of the film leads to a reduction in leakage at equivalent CET.

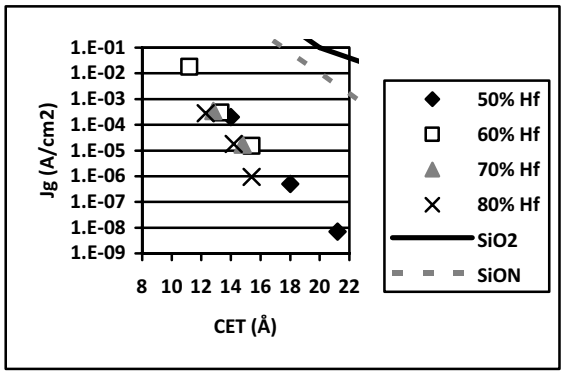


Figure 12. Electrical data obtained from a device employing co-injected ALD $Hf_xSi_{1-x}O_2$ gate dielectric demonstrating low Jg and CET over a wide range of Hf compositions.

Conclusions

It has been demonstrated that through control of several process parameters including process temperature, precursor temperature, process pressure, Ar carrier gas flow rate and precursor pulse time it is possible to obtain a controlled Hf concentration in $Hf_xSi_{1-x}O_2$ films deposited by co-injection ALD. All of the different composition films exhibit similar linear growth characteristics indicative of a true ALD process. The Hf content of the films can be accurately measured using a VUV technique, both for bulk measurements on thick films and for thin films where there is a dilution effect due to the interface layer.

Increasing the Hf content has been shown to reduce the low voltage electrical leakage and increase the dielectric constant of the $Hf_xSi_{1-x}O_2$ films. After integration as a gate dielectric in CMOS device this effect has been proven to reduce the gate leakage current at equivalent CET.

References

1. S. Thomson, P. Packan, M. Bohr, *MOS Scaling - Transistor Challenges for the 21st Century*, Intel Technology Journal Q3'98
2. Inumiya, S. Kamiyama, K. Nakamura, *Hf-based high-k gate dielectrics - Scalability for hp45 node and beyond*, p 132-135, 2006 International Workshop on Nano CMOS, Jan. 30 - Feb. 1 2006
3. I. Burki, C. Rivas, J. Hurst, M. Weldon, H. Yeung, J. Price, P. Lysaght, P.Y. Hung, R. Jammy, *Inline Optical Spectroscopy for Advanced Gate Stacks*, Semiconductor International, 5/1/2008

Ex. 1016, Page 602

ECS Transactions, 18 (1) 575-580 (2009)
10.1149/1.3096504 © The Electrochemical Society

Thin Oxidation Study With Various Thermodynamic Parameters

Ping Linda Zhang [a], Xu Feng Zhang [a], Ting Dun Wen [a], Li Feng Xu[b], Long Xing Shi[b], Stephen E. Webber[c], C. Grant Willson[c], Kim Dean[d]

[a] Department of Physics, North University of China, Taiyuan, Shanxi 030051, China
[b] College of Integrated Circuits, Southeast University, Nanjing, Jiangsu 210096, China
[c] Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, USA
[d] SEMATECH, Austin, Texas 78741, USA

Abstract

The structure analysis of extremely thin thermally grown SiO_2 (< 25 Å) indicates that most silicon atoms share two of the four tetrahedrally-coordinated oxygen atoms with the adjacent silicon atoms. Molecular mechanics simulation of the single chain of SiO_2 tetrahedrons is presented. The extremely thin oxide film is composed of the close-packed short chains of the tetrahedrons perpendicular to Si substrate. Diffusion experiment of an acid vapor through non-oxide polyhydroxystyrene films de-convolutes four prototropic states of fluoresceinamine monolayer to resemble the change of the intermediate states of Si atoms during the thermal oxygen diffusion processes at Si/SiO_2 interface.

Introduction

Thermal oxidation of the silicon wafer is one of the most crucial process steps for IC manufacturing. To obtain a dielectric layer with superb quality, people have tried many different ways. Some people modify the very top silicon surface with the lapping process or other surface modification techniques (1-4); some use the sacrificial oxidation process to remove the native oxide and the initial defects to re-initialize the oxidation process (5); some decrease the oxygen solubility in the silicon substrate (1-7), or even to introduce hydrogen (7) during the crystal growth process to replace the molecular oxygen at the interstitial sites. In general, a post-anneal process after oxidation can help to cure 80% of the OSFs (oxidation-induced stacking faults). People cure the remaining OSF defect by adding ~3% of HCl during oxidation (2). The use of HCl suppresses the formation of the stacking faults. With more stringent requirement due to device scaling, people have restudied the native oxide, which at 20 Å is an effective protection layer for the substrate at the room temperature. People have also revisited the theoretical oxidation models in great details. Deal-Grove model describes the thermal oxidation process at high temperature for thick oxide (>300 Å) well (8).

In this work, our structure analysis of the thermally grown thin oxide at Si/SiO_2 interface indicates that most silicon atoms share two of the four tetrahedrally-coordinated oxygen atoms with the adjacent silicon atoms. Based on six tetrahedrons of SiO_2, the short chains of silicon dioxide tetrahedrons are close-packed and perpendicular to Si substrate. We also performed the diffusion experiment on a series of the amorphous non-oxide (poly-hydroxystyrene) films with several fixed film thickness at room temperature. In-situ detection of the fluorescence intensity affected by the acid vapor diffusion is enabled (9). The empirical diffusion curves through the non-oxide film with various film thicknesses are analyzed and compared with the models. Similar to the existence of the intermediate oxidation states for silicon atom (10-15), the existence of

four prototropic states for the pH sensor is de-convoluted by the diffusion experiments. Single crystal silicon is relatively stable at the room temperature but very chemically active at the elevated temperature (10). Some thermal parameters related to the chemical reaction are also presented.

Experimental

The sample chamber for trifluoroacetic (TFA) acid, (also named as "the acid") has a quartz surface. The acid vapor concentration is 5.366×10^{-3}mol/L at 20°C. On the quartz surface, there is the monolayer pH sensor which is covalently bonded onto the quartz substrate. This single layer pH sensor contains the following: fluoresceinamine (FA) is immobilized on poly(acryloyl chloride) (PACl) polymer while PACl is covalently bonded onto the quartz surface (9). Poly-hydroxystyrene (PHS, also named as the macromolecule or the polymer) in propylene glycol monomethyl ether acetate (PGMEA) is then spin-coated onto this substrate. PHS film is formed with the thickness of 565nm, 965nm, 1.4μm and 4.8μm, respectively. The film is softbaked at 90°C for 60 seconds. Figure 1 is the schematic of the experimental setting. The excitation wavelength is 466nm and the emission wavelength is 516nm for the fluorescence study.

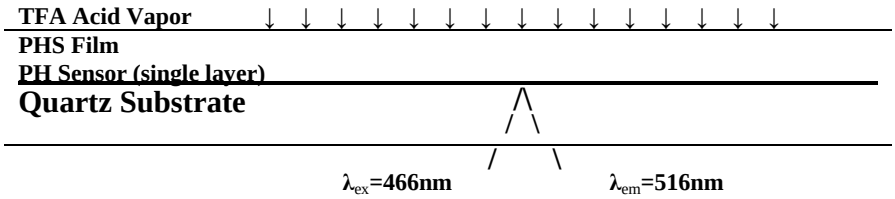


Figure 1. Schematic of the experimental geometry.

Results and Discussion

I. Structural Study of the Interface

a. Bond Angles and Bond Length between Tetrahedrons

The structure of SiO_2 tetrahedrons is unchanged (ignoring the vibration) with variation of the bond angle between the adjacent tetrahedrons. In all of the silicon dioxide film, the sharing of oxygen atoms between tetrahedral is by an apex or vertex only (10-14). No cases are known in which edges or faces of the tetrahedrons are shared. Fused silica is amorphous with disordering of the tetrahedral so that no long-range order exists. But the orientational distribution order of the oxide may exist due to its thermal growth and annealing process, independent of Si(111) or Si(100) substrate orientation.

We speculate that most silicon atoms share two of four tetrahedrally coordinated oxygen atoms with adjacent silicon atoms for the thermal grown film. At the linear oxidation, it forms a series of closed-packed short chains of the tetrahedrons at the interface, perpendicular to the substrate.

Theoretically when a vertex is shared between two tetrahedral, the maximum value of the Si-Si separation is two Si-O bond length for collinear Si-O-Si bonds, but a considerable range of Si-O-Si angles (102°-180°) is possible (10-14), consistent with the distance between O atoms of different tetrahedral being not less than the edge-length of the tetrahedron. The wide distribution of the angles is mostly in the range of 130°-150°.

We conducted the simulations based on the molecular mechanics modeling method which uses the law of classical physics. Energy minimization (or the gradient of the energy) of six silicon dioxide tetrahedrons is thoroughly studied (Figure 2). The bond

Ex. 1016, Page 604

angle distribution of 150°-170° is obtained. In the single chain simulation, we ignored the packing potential. Table 1 reports our study. Experimentally the neutron scattering and neutron diffraction is possible to confirm the conformation of the tetrahedrons of the thick silicon dioxide film (after a long time thermal oxidation at high temperature). In Figure 2, the rotational relaxation of six tetrahedrons along the axis perpendicular to the interface is preferred. (e), (f) and (g) represents one of millions of possible rotational relaxation moments of (a) where their tetrahedron secondary orientation is collective. Most likely (a) represents the linear growth natural oxide formed by the elongated short chains of silicon dioxide tetrahedra.

Table 1 Bond Angles and Bond Length between Tetrahedrons

Tetrahedron Bond Length	**Value**	**Bond Length btw. Tetrahedrons**	**Value**
O-O	0.262nm		
Si-O	0.162nm	Si-Si	0.31nm, vary
Tetrahedron Bond Angle	**Value**	**Bond Angles btw. Tetrahedrons**	**Value**
<O-Si-O	109.5°	<Si-O-Si	150°-170°, vary

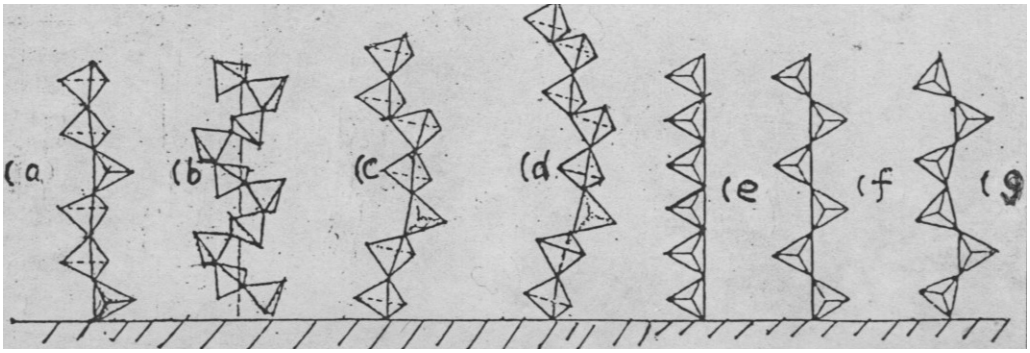


Figure 2(a)(b)(c)(d)(e)(f)(g). Two-dimensional projections of three-dimensional plots of six tetrahedrons. Single chain mostly formed from six SiO_2 tetrahedra, assuming that the tetrahedron chains are perpendicular to the interface. Most likely (a) represents the linear growth natural oxide formed by the elongated short chains of oxide tetrahedra.

At the interface, such a bond arrangement is forced upon the silicon atoms by the fact that they have to bond to silicon on one side of the interface and oxygen on the other. The dangling bond of Si atoms is also located at the interface. The distribution of the oxidation states and the orientational order of the tetrahedron chains can change with the preparation conditions and with the quality of the interface. Thin oxide could be extremely close-packed, short chains of several oxide tetrahedrons on Si substrate at the early stage. The requirement for a good oxide film to be useful is that the film is either single crystal or amorphous.

Pauling has listed a set of rules based on an ionic model (14). Although no one now accepts a purely electrostatic model for silicon dioxide, Pauling's rules are still reasonably accurate as long as the partial charges on the atoms are sufficient to make electrostatic repulsion militate against the sharing of edges or faces by tetrahedra since this places the positive centers too near to each other.

b. Electronic and Vibrational Property

Experimentally, Si+, Si++, Si+++, the intermediate oxidation state 1+， 2+， 3+, are found at Si/SiO_2 interface (15), which are not stable in normal silicon compounds. These four intermediate states are usually convoluted and coexisted for most detection measurement (2-3, 5, 8, 15). Please notice that the intermediate states mentioned above or later about our pH sensor is different from the ionization energy of silicon atom, which

Ex. 1016, Page 605

has the first ionization and second ionization energy, such as Si^{+1}, Si^{+2} (Table 2). In the interface diffusion experiment we reported, however, we are also able to separate and detect four prototropic states of the pH sensitive dye at macromolecule/SiO_2 interface, analogous to Dr. Himpsel's work. The intermediate prototropic states of the monolayer with cation, neutral, monoanion, dianion at different pH are observed by the de-convolution by dispersing the time of the events through the fluorescence experiment.

TABLE 2. Exothermic Reaction and Ionization Energy of Si and IR Spectrum of SiO_2 at room Temperature

Energy State	Value (ref)	Vibration Band(cm^{-1})	"Thermal" Temperature
Exothermic	ΔH=-8.727eV/molecule	1126	"1621K"
Si^{+1}	8.15eV	809	"1165K"
Si^{+2}	16.34eV	452	"651K"
Si^{+3}	33.49eV	Thermal expansion($x10^{-6}$)	0.6
Melting point	~1700°C	Therm. Conductivity(W/cm-°C)	0.014

By observing the energy shifts of the silicon 2p level induced by the bonding partners, it is possible to reveal the different bonding configurations at the interface. Silicon itself is chemically stable at the room temperature and is chemically active at the elevated temperature (10). The higher the temperature for the reaction, the farther apart the atoms and molecules due to the increased vibration energy.

II. Phenomenological Study of the Interface

Thermal oxidation of Si(100) or Si(111) is the best example for the diffusion study. That is why Deal-Grove model can accurately describe oxygen diffusion process in the oxide > 300 Å at high temperature, the diffusion-controlled thermal oxidation. Many scientists and engineers contribute their work to the early stage oxidation study (1-3, 5, 8, 15-16). Theoretically it is hard to accurately describe the early stage thin oxidation (<100Å). Mathematically it is a moving boundary diffusion modeling problem – the boundary of silicon and silicon dioxide is constantly changing. Experimentally most detection method is not sensitive enough to detect this weak signal. Chemically it is the reaction-rate controlled, on-going fast exothermic reaction for silicon and oxygen at high temperature. In thermal oxidation the chemical reaction between oxygen and silicon and the diffusion of oxygen in silicon dioxide film occur simultaneously at linear oxidation regime.

a. De-convolution Experiment

At the early stage of oxygen diffusion, the silicon-oxygen chemical reaction at various stage of four "electronic" states of silicon atoms are convoluted. The signal from just one or two layer of Si atoms is next to impossible to detect. To design an experiment to de-convolute the reaction-rate controlled diffusion from the reaction, to distinguish or to further de-convolutes the intermediate states of silicon atom, a mimic experiment, that is, a fixed boundary diffusion experiment of the acid vapor in PHS film is performed. Fluorescent monolayer, a single layer of the fluorescent dye molecules, enables the detection. Here the acid vapor mimics O_2 or HCl molecules while PHS film mimics the oxide film. The interaction between the acid vapor and the monolayer pH sensor mimics the interaction between oxygen molecules and silicon atoms. This is one of few experiment available to de-convolute the early stage, linear response, reaction-rate controlled diffusion process. Various PHS film thickness further disperses or de-convolutes four prototropic states of the monolayer during the acid diffusion. Figure 3 provides four prototropic states of the fluorescent monolayer with respect to pH, which is

Ex. 1016, Page 606

equivalent to the acid vapor diffusion time. With respect to the reaction, the transition from the dianion to monoanion for the monolayer in Region I happens instantaneously, independent of the existence of the diffusion barrier, even for the thick PHS film. Different initial monolayer condition may influence where the diffusion curve starts from Region I. It is hard to normalize the diffusion curves from Region I. Experimentally we monitored and recorded the diffusion of the acid in PHS film.

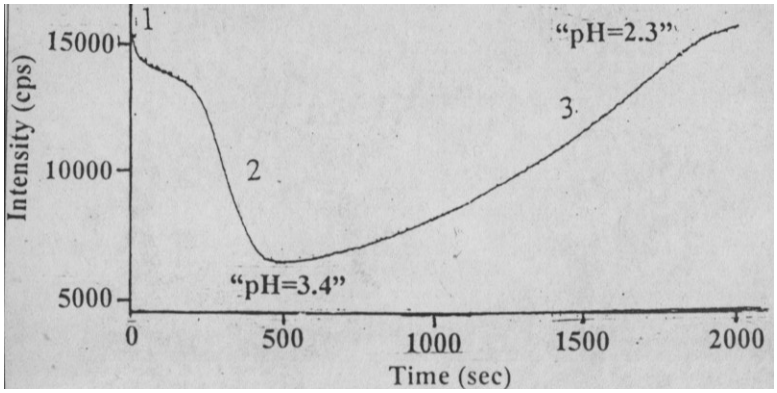


Figure 3. The fluorescence intensity of pH sensor versus time (assuming constant acid mass increase through diffusion). Region I represents the transition from dianion to anion; Region II represents the transition from anion to neutral; Region III represents the transition from neutral to cation.

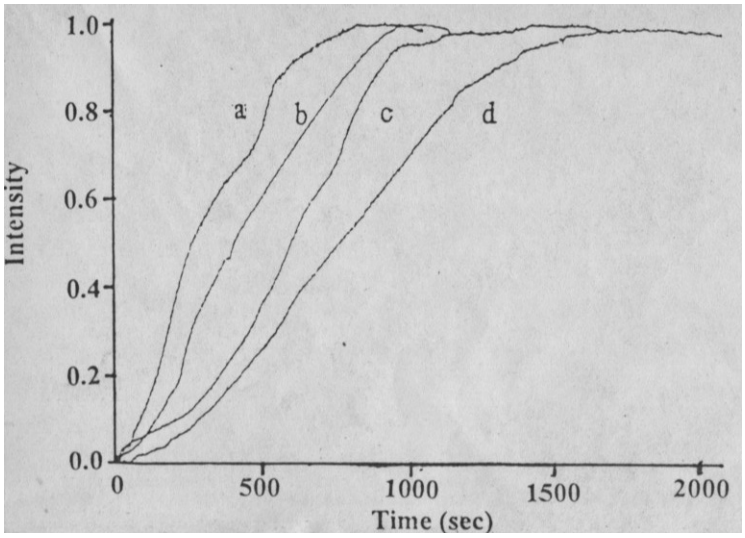


Figure 4. Normalized fluorescence intensity of the pH sensors with respect to the diffusion time. a, b, c and d represent the films with the thickness of 565nm, 965nm, 1.4μm and 4.8μm, respectively.

Figure 4 provides Region III diffusion curves a, b, c and d with respect to the film thickness of 565nm, 965nm,1.4μm and 4.8μm. In Figure 4, the diffusion time is re-initialized to zero for the fluorescence minimum and the fluorescence maximum is used to normalize the diffusion curves that are leveled later. This provides the study for further understanding of the diffusion of oxygen in the reaction-rate controlled linear regime in silicon dioxide.

b. Discussion of De-convolution Experiment

Ex. 1016, Page 607

In Figure 3, the acid vapor diffusion in Region I is instantaneous. This supports the pinhole theory (16). The "short shoulder plateau" between Region I and Region II show that a gradual uniform change of the global film happens before it participate at the second stage of the prototropic transition. This prevents further application of the pinhole theory. The reason we studied the fluorescence from cation is that it is possibly the only emission species and the neutral form does not fluorescence (the quantum yield of the neutral form is 0.00 while the quantum yield of the cation is 0.39). Various diffusion models, such as the error function, have been tried to fit to the empirical data in Figure 4.

The bridging oxygen site between two polyhedra (2) is the same as the shared vertex (oxygen) for the tetrahedrons in this paper. The bridging oxygen or sharing vertex of silicon dioxide tetrahedrons is essential to the dielectric property of the oxide. The larger the fraction of bridging to non-bridging sites, the most cohesive and less prone to damage the oxide is.

Conclusions

We represented our thin oxide structure study and our diffusion data detected by fluorescent monolayer. The new experiment and the new theory developed exclusively for ultra thin oxide is in great demand.

References

1. J. G. Wilkes, *Silicon Technology*, in K.A. Jackson *Processing of Semiconductors in Materials Science and Technology*, edited by R. W. Cahn, P. Haasen, E. J. Kramer, Vol. 16, VCH-Verlag, (1996); Science Press, (1999), (in Chinese).
2. S.A. Campbell, *The Science and Engineering of Microelectronic Fabrication*, 2nd Edition, Chapter **4** (references therein), Oxford University Press, (2001).
3. M. Quirk, J. Serda, *Semiconductor Manufacturing Technology*, Chapter **10** (references therein), Prentice Hall, (2001).
4. J. Z, Zhang, *Chemical Mechanical Polishing Technology of Single Crystal Silicon Wafers*, Chapter **5**, Chemical Industry Press, 2005, (in Chinese).
5. H. Xiao, *Introduction to Semiconductor manufacturing*, Chapter **5**, Prentice Hall, (2001).
6. F. Shimura, *Semiconductor Silicon Crystal Technology*, Academic Press, San Diego, 1989.
7. T. Hattori, K. Takahashi, H. Nohira, T. Ohmi, in *Semiconductor Technology (ISTC 2001)*, edited by M. Yang, Vol. **1**, 169, ECS Asia, (2001).
8. B. E. Deal, A. S. Grove, *J. Applied Phys.*, **36**, 3770 (1965).
9. P.L. Zhang, A. Eckert, G. Willson, S. Webber, J. Byers, *SPIE*, **3049,** 898, (1997).
10. R. He, X. Fang, S. Niu, *Carbon, Silicon, Germanium*, Chapter **10**, in Inorganic Chemistry Series, Vol. **3**, Science Press, (1988), (in Chinese).
11. A. F. Well, *Structural Inorganic Chemistry*, *5th Edition*, Chapter **23**, Oxford University Press, (1984).
12. J. E. Huheey, *Inorganic Chem.*, 2rd Edition, Chapter **14**, Harper and Row, (1978).
13. A. Dai, J. Yi, Z. Yan, Q. Zhang, *Course of Inorganic Chemistry*, Chapter **16**, People Education Press, (1958), (in Chinese).
14. L. Pauling, *The Nature of the Chemical Bond*, 3rd ed., p544, Cornell University Press, Ithaca, N. Y., (1960).
15. F. J. Himpsel, F. R. Mcfeely, A. Talab-Ibrahimi, J. A. Yarmoff, G. Hollinger, *Phys. Review*, **B**, **38**, 6084, (1988).
16. A. G. Revesz, R. J. Evans, *J. Phys. Chem. Solids*, **30**, 551, (1969).

Ex. 1016, Page 608

ECS Transactions, 18 (1) 581-586 (2009)
10.1149/1.3096505 © The Electrochemical Society

Solving Arcing Issues in CVD Processes

Evan Yan[a], Jerry Xu[a], Xianyuan Li[a], James Chen[a], Yingjie Chen[a], Dae Joo Cho[b], Lanchia Chan[c], Lily Ji[c] , Deenesh Padhi[c] , and Bob Zhang[d]

[a]Applied Materials, Shanghai, China
[b]Applied Materials, Icheon, Korea
[c]Applied Materials, Santa Clara, CA, USA
[d]Wuhan Xinxin Semiconductor Manufacturing Corporation, Wuhan, China

Abstract

As the semiconductor industry continues to scale feature sizes, higher deposition/etch rate and better gap-fill abilities are continuously to be required for chemical vapor deposition (CVD) and other similar processes. To achieve higher deposition rates, higher plasma power and density are required, which can lead to plasma arcing. Arcing is a severe issue common to CVD processing. Minor arcing can generate particles/defects. Severe arcing can completely destroy wafers and reactor components. In this paper, three typical arcing cases are examined using a 300mm CVD system. These include the chamber body arcing using a high density plasma (HDP) phosphosilicate glass (PSG) chamber, the electrostatic chuck (ESC) arcing using an HDP undoped silicate glass (USG) chamber, and the wafer arcing with an advanced patterning film (APF) chamber. In each case, root causes were established and the corresponding solutions to prevent arcing were demonstrated. Brief theories of arcing in CVD processes and arcing detection methods are also discussed.

Introduction

Arcing issues in semiconductor manufacturing processes occur commonly at low pressure, high plasma power and high plasma density. The problem becomes more severe for nanoscale devices because they are more susceptible to arc-induced damages. For CVD tools, the common issues include wafer-level and ESC arcing in HDP chambers, and wafer-level arcing in plasma enhanced CVD (PECVD) chambers. These arcing issues can cause physical damage to chamber components, chamber contamination, and RF system failure. For device wafers, arcing problems can cause dielectric breakdown, interconnect open, short circuit failures, contamination, and gate oxide integrity (GOI) failure, etc. Understanding and correcting the root causes of arcing is crucial to avoid reduced mean time between failure (MTBF) of CVD chambers and low up-time, which can impact the overall fab efficiency.

Ex. 1016, Page 609

Case Study I. Chamber-body-induced arcing in new HDP-PSG chamber

Problem Description: An arcing issue was identified in an HDP-PSG chamber during the process qualification. The RF power was 1800W at a chamber pressure of 4.3mTorr. Through the viewport, the plasma was observed to be flickering with visible arcing during the process. At the same time, the chamber pressure fluctuated violently, preventing the chamber pressure from reaching the pressure setpoint within the process time window. Also, the RF reflected power was much higher than normal. After opening the chamber, white arcing marks (Fig. 1) around the chamber wall were identified. The arcing marks were easy to remove, but the arcing caused chamber body erosion and poor particle performance.



Fig. 1 PSG chamber arcing damage

Root cause analysis: We have identified four potential root causes. First of all, chamber surface roughness is high for the new chamber and has some sharp points prone to plasma discharge. Past experience has shown that no further chamber wall arcing occurs once the chamber has been used in production for more than a few thousand wafers. Secondly, unstable plasma occurred during the pressure transition between two process steps. The pressure fluctuation was caused by incorrect PID pressure control constants for the production recipe. However, PID tuning is limited therefore absolute throttle valve position was used to set chamber pressure for the recipe step transition, instead of relying on the pressure servo function alone. Recipe optimization included optimizing PID pressure control to ensure the process pressure (>4mTorr) during deposition and minimize pressure variation using step control at “heat up”, “chuck on” and “pre-bias” steps. Thirdly, the top/side RF matching network was not properly tuned to minimize reflective power. A customized soft landing impedance trajectory was introduced to supplement the standard chamber start up procedure, significantly reducing the reflected power to <150W. Lastly, there were H_2O and organic gas residues in the chamber due to improper chamber conditioning. These residues would start to outgas during the high power HDP process. Those gases were easily ionized by plasma and led to plasma discharge or arcing. The solution was to implement optimized baking and aging procedure for these new chambers. In summary, the arcing was solved by implementing a combination of chamber surface polishing and conditioning, tuning RF top/side reflect power, and recipe optimization.

Case Study II. ESC arcing in an HDP-USG chamber

Problem Description: While running a dummy wafer, ESC arcing occurred in an HDP-USG chamber (wafer count less than 3000) and was accompanied by two alarms - ESC current fault and high IHC helium leak rate. During a step change in bias power, the chucking force became unstable and the backside helium pressure caused the wafer to lift from the ESC surface. The arcing occurred after ESC current fault or high IHC leak rate, especially for the 1st wafer processed after chamber cleaning.

Ex. 1016, Page 610

Root cause analysis: The root cause was determined to be a combination of two problems areas. Wafer-related problems, such as high backside particles on the dummy wafer, wafer bowing, and incorrect wafer handoff position, can create poor wafer-to-ESC

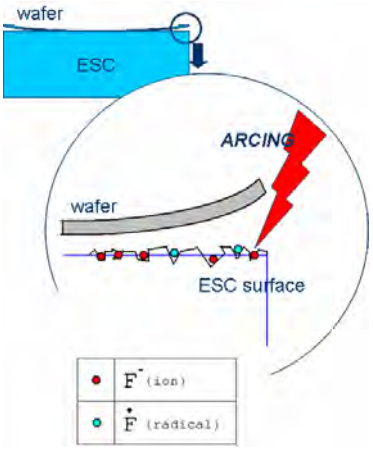


Fig. 2 Sketch of arcing mechanism

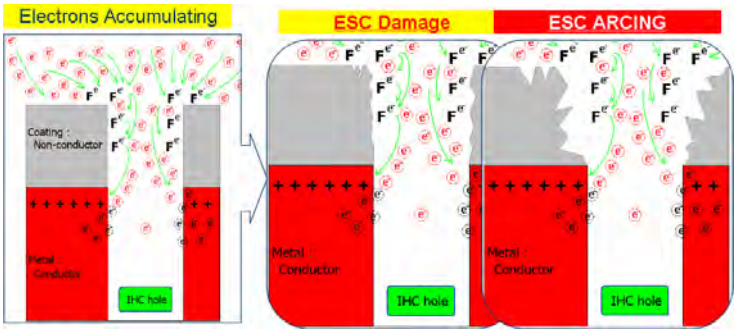


Fig. 3 Mechanism of high leakage current and ESC damage

contact. When an already highly stressed wafer receives more compressive stress during HDP-CVD deposition, the wafer bowing is increased. As a result, a gap between the wafer and the ESC is developed, which allows residual fluorine ions and radical residues from the *in-situ* chamber clean process to damage the ESC protective insulation, making chucking force unstable during RF bias power changes. The fluorine concentration is higher with high NF_3 flow and for the first wafer after cleaning. The damaged insulation allows arcing to occur due to plasma leak under the wafer in high bias RF power process (Figs. 2 and 3). Excess fluorine ions near the backside helium channels, which are the lowest resistance ground path, can work as electron messengers and generate excess electrons, leading to higher temperature, ESC damage, and arcing.

The arcing was mitigated by using better quality dummy wafers and implementation of optimized clean recipes. The clean recipe was modified to reduce the fluorine density by reducing the cleaning time and adding fluorine scavenge step using H_2. In addition, a bake recipe was created to accelerate the fluorine residue removal rate.

Ex. 1016, Page 611

Case Study III: Wafer arcing in an APF chamber

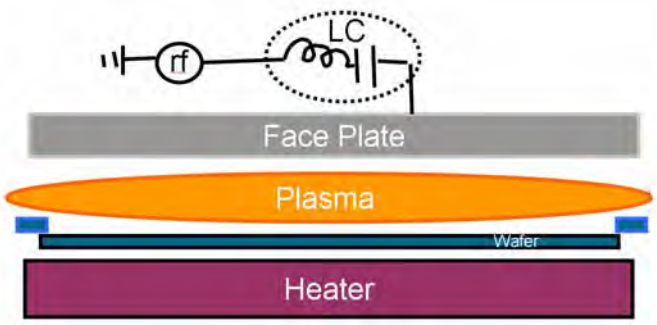


Fig. 4 Diagram of PECVD chamber

In general, possible arcing sources in a PECVD reactor include:

1. Impedance mismatching which leads to plasma voltage insatiability.
2. High DC bias, plasma and non-uniform plasma formed by special RF frequency, pressure, spacing, chemistry and electrode.
3. Wafer ground path (bevel deposition of conductive materials), wafer flatness, materials electrical properties and surface roughness.
4. RF ground path and stability, heater to wafer contact, and RF ground mesh diameter.
5. Shadow ring overlap and gap confines plasma at wafer edge (in shadow ring process).
6. Face plate surface finish, sharp gas holes, whole angles, and pressure drop.

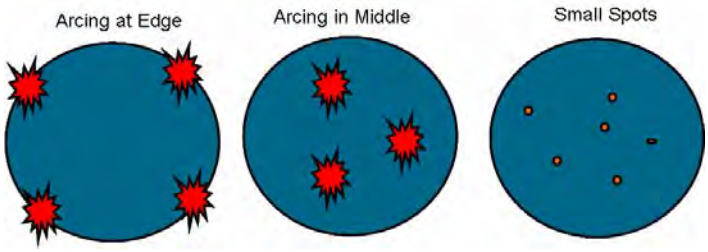


Fig. 5 Three different arcing patterns

Problem Description: APF arcing was observed on wafers with an APF/W filmstack. There were three different types of arcing patterns, with arc-induced damage observed at the edge, the middle and in small areas across the wafer (Fig. 5).

Root cause analysis: Arcing at the edge may be caused by shadow ring coverage and bevel contamination. Backside deposition of conductive materials such as TiN or W can provide alternative ground paths (Fig. 6). This will result in a change in impedance and instability in power delivery resulting in arcing. For arcing in the middle of the wafer, the root cause may be wafer contamination, surface roughness, or dielectric materials failure above the W layer. If the W surface is too rough, the sharp peaks can act as field concentrators and thus lead to local plasma imbalance. The third pattern shows small arcing spots across the wafer. Arcing can occur if a local voltage gradient is established

Ex. 1016, Page 612

on the wafer surface. It may be caused by ring-type defects and wafer contamination. This can be caused by excessive deposition/ashing cycles resulting in damaged dielectric surface and non-uniform plasma across the wafer (Fig. 6).

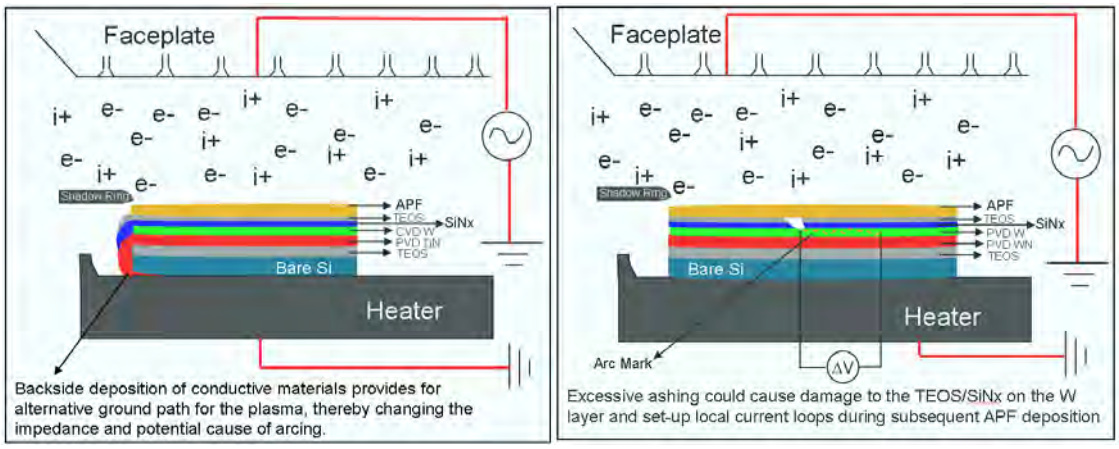


Fig. 6 Arcing mechanism caused by bevel contamination and dielectric failure

Arc Detectors and Systems

Real-time, high-speed analyses and detects of plasma arcing are very important for reducing wafer loss and improving yields. Advanced RF detectors provides real-time, high-resolution analyses of plasma micro-arcing which can cause damage to the target, the film being deposited, and even the wafer surface in CVD and ETCH processes. Non-intrusive detectors which connect directly to the reactor's RF delivery system to collect voltage and current information are particularly attractive. High speed DC Bias monitoring equipment can be useful to provide a high-resolution picture of plasma condition and offer arc detection and prevention feedback. The limitation of these systems is their inability to detect micro-discharges between plasma sheath and wafer. The optical emission spectrometers (OES) often used as endpoint systems can be used to detects plasma anomalies by detecting instable optical wavelength emission. Again, OES systems are non-invasive and can collect spectral information through any convenient viewport. OES systems may not, however, have enough sampling speed to detect arcing which happens in the microsecond range.

Conclusion

Plasma arcing in HDP-CVD and PECVD reactors can cause yield loss and reduced fab efficiency due to unscheduled downtime. However, by understanding the root cause of arcing in a particular application, steps can be taken to reduce or eliminate this problem. We have presented three case studies where this methodology was used to solve serious arcing problems in high-volume 300mm deposition systems.

Ex. 1016, Page 613

Acknowledgments

Special thanks go to Shengyu Jiang, and Many thanks to the hardware and process engineers from CVD teams at Applied Materials and WXIC, without whom this paper would not have been possible.

ECS Transactions, 18 (1) 587-591 (2009)
10.1149/1.3096506 © The Electrochemical Society

Study on Bit Line $TiCl_4$ TIN Process Optimization for Stack DRAM IDDS Standby Failure Improvement

Liang Chen, Zhi-Chao Li, Xing-Hua Song, Lei Xu, Yi-Hui Lin,
Hsueh-Hung Wei, and Paul-Chang Lin

SMIC (Semiconductor Manufacturing International Corp),
18 Zhangjiang Road, 201203, Shanghai, China

Dynamic Random Access Memory (DRAM) plays an important role in semiconductor industry. For 0.13um stack DRAM (D13) process, Bit line (BL) is a critical loop and acts as a contact between active area (AA) and periphery area. Normally, BL consists of BL glue layer Ti/TiN and BL W and $TiCl_4$ has been used for D13 BL glue layer Ti/TiN deposition. D13 suffers electrical drain-to-drain standby (IDDS) failure, meaning junction electrical leakage. It has been found that BL glue Ti/TiN has strong influence on AA Ti diffuse. In addition, IDDS standby failure is enhanced with deeper TiSix formation. In this study, we present the effect of $TiCl_4$ glue TiN film's quality on D13 IDDS standby failure.

INTRODUCTION

Stack DRAM plays an important role in semiconductor industry. As devices shrinking into below 0.13um and contact height/diameter ratio increase, $TiCl_4$ chemistry, with high step coverage and in-situ contact TiSix formation, has been adopted in advanced semiconductor manufacturing for contact glue layer Ti/TiN. In $TiCl_4$ chemistry, Ti is deposited by the reaction between $TiCl_4$ and H_2 with RF-plasma enhancement. Ti film reacts with active area (AA) to form TiSix in-situ at high temperature, which can eliminate an RTP step, whereas $TiCl_4$ reaction with NH_3 forms a TiN film. Recently we have found that D13 suffered from an IDDS standby failure. The corresponding TEM pictures showed excessive TiSix was formed at the bottom of contact, which consequently induced junction electrical leakage. According to the study from H.S. Park, enhanced post-nitridation in PECVD-Ti was recognized to reduce the excessive formation of TiSix after high temperature post annealing [1]. In this study, a detail study on TiN process optimization to prevent IDDS standby failure is presented.

Ex. 1016, Page 615

EXPERIMENT

The experiment was executed on an AMAT Centura5200 $TiCl_4$ system, which has two main process chambers, PECVD Ti chamber and CVD TiN chamber, as shown in Fig. 1 below.

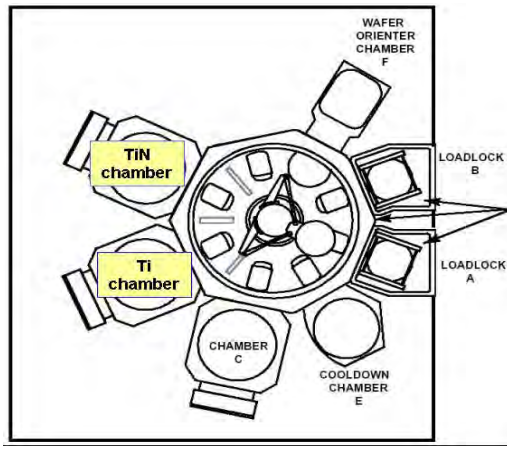


Fig. 1An AMAT Centura 5200 TiCl4 system [2]

For BL glue layer Ti/TiN, AMAT Centura 5200 $TiCl_4$ was processed after pre oxide WET clean. At first Ti film was deposited by the reaction between $TiCl_4$ and H_2 in PECVD chamber at 600~700℃. Then TiN was processed on the Ti surface through the reaction between $TiCl_4$ and NH_3 in CVD chamber at 600~700℃. The standard Ti/TiN deposition chemical reactions are as follows [2]:

$$TiCl_4 + H_2 \rightarrow \mathbf{Ti} + 4HCl$$

$$6TiCl_4 + 8NH_3 \rightarrow 6\mathbf{TiN} + 24HCl + N_2$$

If the ratio of $TiCl_4$ to NH_3 were kept at a proper value of 6:8, the component of N would be equal to Ti. Otherwise, the reaction would result in an unequal Ti/N ratio film. Such reaction would produce N rich high resistivity TiN film if the ratio of $TiCl_4$ to NH_3 were less than 6:8 [2]. And on the contrary, Ti rich low resistivity TiN film will be formed. The split condition was focused on the content of N and TiN film stack. Firstly, TiN's N ratio of constituent is controlled through adjusting the flow ratio between $TiCl_4$ and NH_3 to form N rich TiN or Ti rich TiN; Secondly, we designed 2 steps TiN deposition with different Ti composition. Then, all splits' TiSix thickness and IDDS standby failure were investigated to verify 2 factors' effect.

RESULTS AND DISCUSSION

TEM shown in Fig. 2 shows IDDS failed sites have deeper TiSix formation, which leads

Ex. 1016, Page 616

to junction electrical leakage.

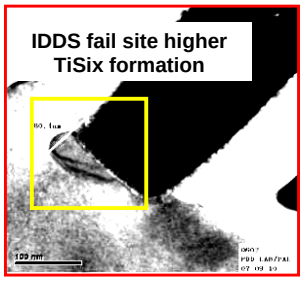


Fig. 2 IDDS fail site TEM

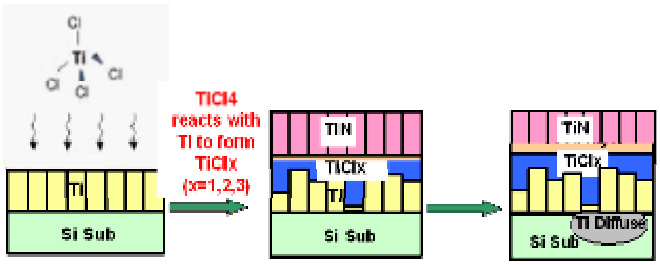


Fig.3 Higher TiSix formation mechanism

As shown in Fig. 3, during normal TiN deposition, TiCl4 would react with Ti film to form TiClx (x=1,2,3), which can lead to deeper TiSix formation. As mentioned eralier, an enhanced Ti post-nitridation can reduce the formation of excessive TiSix. In our experiments, TiN effect on junction leakage and IDDS standby failure is discussed.

$TiCl_4/NH_3$ ratio effect

By controlling the amount and the ratio of $TiCl_4/NH_3$, two different qualities of TiN were obtained, as shown in Table 1. It is worth noting that this ratio was kept between X_1 and X_2 satisfying the relationship $X_2 < 6/8 < X_1$.

Table 1: $TiCl_4/NH_3$ split for TiN component

Condition	$TiCl_4/NH_3$ Ratio
1	X_1
2	X_2
Remark: $X_2<6/8<X_1$	

Under condition 1, TiN was Ti rich since excessive reactant $TiCl_4$ was applied. In contrast, condition 2 render an N rich TiN film since NH_3 was excessive.

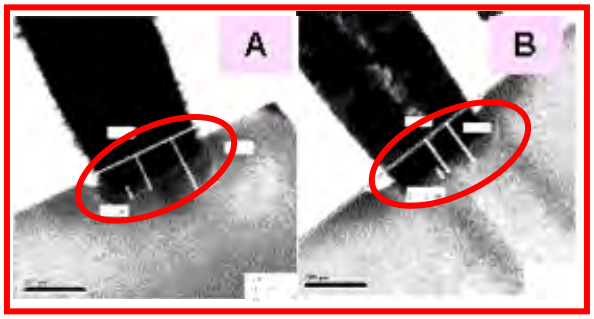


Fig. 4: TEM image of TiSix formation: A) $TiCl_4/NH_3$ ratio X_1; B) $TiCl_4/NH_3$ ratio X_2

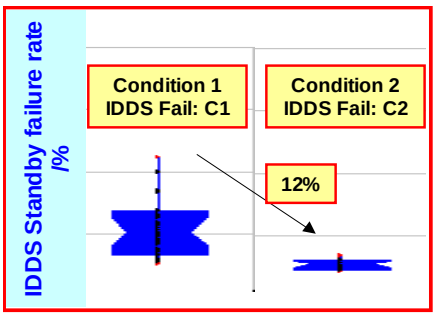


Fig. 5: IDDS performance vs different TiN composition

Figure 4 shows that the TiSix was very deep under the bottom of BL contact at baseline condition X_1. After decreasing the $TiCl_4/NH_3$ ratio to X_2, TiSix deeper formation

disappeared. Fig. 5 shows that the baseline condition suffered serious IDDS standby failure, as evident from the fail rate reach C_1. But under the condition2, IDDS standby failure rate decreased significantly to C_2. Total fail ratio improved by 12%. Under condition2, the N rich TiN film could form a barrier layer, which could prevent further TiSix formation and hence IDDS to result in less device junction electrical leakage.

TiN film stack:

As shown in the previous section, an N rich TiN film is beneficial to avoid IDDS standby failure. However, the corresponding recipe, with certain extent of high thickness uniformity issue, is too slow for FAB mass production. A new TiN film stack, shown in Fig. 6, is developed to investigate its effect on TiSix formation and IDDS standby failure. To fabricate this stack, TiN deposition was divided into 2 groups, Dep1 and Dep2. The ratio of $TiCl_4/NH_3$ was X_2 and X_1 for Dep1 and Dep2, respectively. Dep1 produced an N rich TiN with thickness ranging from T_1~T_4 ($0=T_1<T_2<T_3<T_4<$total TiN thickness. TiN from Dep2 is Ti rich. The corresponding details are listed in Table 2.

Table 2: Thickness split for TiN film stack

Condition	Dep1 thickness	TiSix formation thickness	TiN NU%
1	T1	T	U%
2	T2	0.89T	1.31U%
3	T3	0.62T	1.42U%
4	T4	0.50T	2.27U%
Remark: 0=T1<T2<T3<T4<Total thickness			

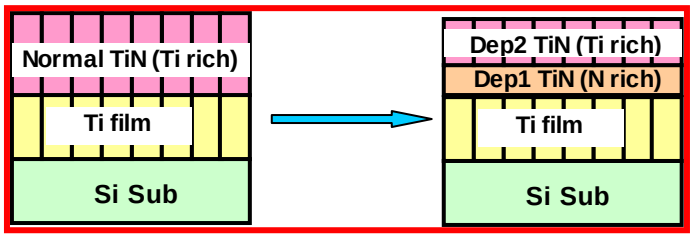


Fig. 6: TiN film stack optimization

All splits (condition 1 to condition 4) suffered from the excessive TiSix formation. However, TiSix formation decreases with increasing Dep1 TiN thickness, as shown in Fig. 7.

Ex. 1016, Page 618

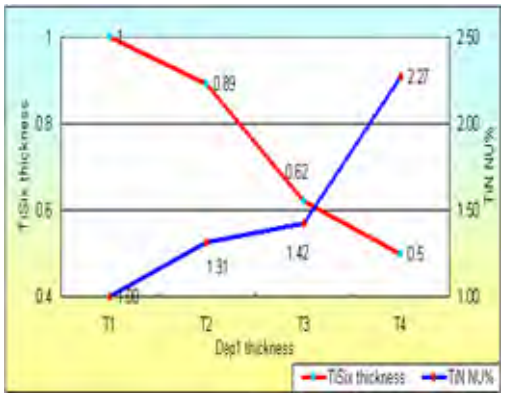


Fig. 7 Correlation between TiSix and Dep1 thickness

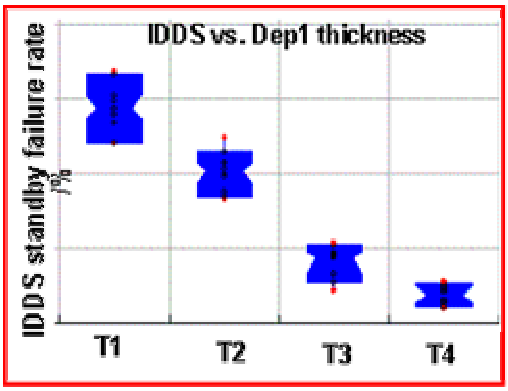


Fig. 8 IDDS standby failure rate vs. Dep1 thickness

As shown in Fig. 8, IDDS standby failure rate decreased rapidly with increasing Dep1 TiN thickness. IDDS failure could be improved by 7% when Dep1 thickness is set at T_3, and a 9% yield would be saved if Dep1 thickness was kept at the value of T_4. However, if Dep1 TiN thickness increased to T_4, total TiN thickness uniformity would trend up and could not be able to meet manufacturing goal. In addition, throughput would be low as Dep1 had slow deposition rate than that of Dep2. Therefore TiN film stack with Dep1 thickness set at T_3 was selected for IDDS standby failure. In short, the optimized TiN process improved IDDS standby failure and reliability simultaneously.

CONCLUSION

D13 IDDS standby failure has been shown to be due to the excessive TiSix formation, leading to junction electrical leakage. The optimization on BL contact glue layer TiN includes increasing N composition and dividing TiN into 2 steps. The results have been validated by a significant yield improvement.

ACKNOWLEDGMENTS

Authors would like to thank their supervisors, Allan_Zhou, Division head of SMIC-S1-TF and Daniel_Fu, Center head of SMIC-S1 center, for their encouragement and guidance.

REFERENCES

[1] Hee Sook Park, Kwang Jin Moon, Myoung Bum Lee, Sang Bom Kang, Gil Heyun Choi, Young Wook Park, Joo Tae Moon, and Hee Sook Park, in: VLSI Technology, Systems, and Applications, 2001. Proceedings of Technical Papers. 2001 International Symposium, on 18-20 April 2001 Page(s): 93 – 96.
[2] Bing Wu et. al., “Applied Materials Technical TiCl4 Ti_TiN_Training”, Version: D; Page(s): 91-102, 20 Nov. 2000.

Ex. 1016, Page 619

ECS Transactions, 18 (1) 593-599 (2009)
10.1149/1.3096507 © The Electrochemical Society

Second Generation Copper Dielectric Barrier (BLOk II): k Reduction and Interface Control To Improve TDDB and EM

L.-Q. Xia, V. Zubkov, W. F. Ye, M. Y. Shek, Z. Cui, and M. Naik

Blanket Dielectric Division, Applied Materials, Inc. Santa Clara, CA 95052, USA

Sub-45nm backend device integration requires material improvements for further RC reduction. One of the key candidates to achieve lower RC is the dielectric barrier. The dielectric constant of the copper barrier can be reduced to <4.0 by additional carbon doping. In addition, this new material can still maintain comparable copper barrier properties and etch selectivity to the current 45nm dielectric barrier baseline. To meet the challenges of the device shrink, advanced interface management is required to achieve good TDDB and EM. An Enhanced Nitride Interface (ENI) concept was adopted to maximize the interfacial mechanical and electrical strength without negative impacts on copper electrical performance.

Introduction

As IC manufacturing migrating to sub 45nm geometries and beyond, device speed is significantly impacted by the capacitance of the insulating materials sandwiched between copper interconnect layers. Insulating materials with lower dielectric constants need be integrated with the copper wiring layers. A porous low k (~ 2.5) material has been successfully integrated into the device structure. Further k reduction can be achieved by modifying the copper barrier dielectric (BLOk). Currently the baseline BLOk for 65nm production is an amorphous dielectric material, composed of SiCN with a k value ~ 5.1. A total of over 6% capacitance reduction can be achieved by reducing the corresponding dielectric constant of this barrier from 5.1 to 4.0. It has been found that adding porosity to the insulating materials is no longer a viable solution because it degrades too much the etch selectivity and the Cu diffusion [1]. Our new approach is to modify the BLOk deposition base chemistry by adding a carbon precursor to achieve a low k value less than 4.0. Carbon, having lower polarizability, can further reduce the dielectric constant. In addition, a C-rich film enhances polymer formation during the etch process to maintain the etch selectivity. No degradation in Cu barrier or other dielectric properties from this new chemistry is observed [2]. BLOk II low k barrier films have since been qualified for 45nm manufacturing. As the bulk low k dielectric becomes more porous, the Cu barriers are becoming crucial for backend of line device reliability. The major challenges for Cu interconnect reliability include Time-Dependent-Dielectric-Breakdown (TDDB) and electro-migration (EM). The interface between BLOk and Cu becomes critical in controlling the corresponding electrical properties and reliability performance [3,4]. Reducing interface impurity has been identified as the key for good adhesion and high breakdown performance. *In-situ* CuO removal using an NH_3 plasma has been adopted since the 90nm device node, but has also been shown to be insufficient for devices with smaller features. Minimizing the C content is another knob for improving the electrical performance. Various interface enhancement techniques, e.g., CuSiN or CoWP, have been evaluated. However, they are both very sensitive to the Cu surface conditions and

difficult to implement in a manufacturing environment. BLOk II utilizes a new initiation scheme to form a controllable interface layer to optimize the interface mechanical and electrical properties without any negative impacts on copper electrical properties.

EXPERIMENTAL SETUP AND RESULTS

Copper Barrier k-value Reduction

Early attempts to decrease the barrier film dielectric constants were carried out by adjusting the plasma density during deposition without changing the corresponding process chemistry. Fig. 1 shows that the k is reduced at the expense of film density. As a result, the film becomes more porous and the dry etch rate of the blanket film increases.

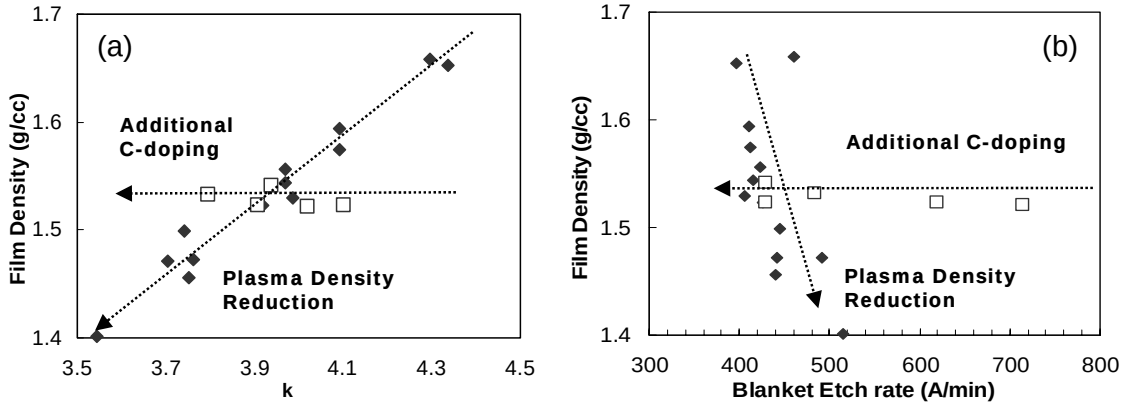


Fig. 1 Effect of plasma density reduction and C-doping on a) copper barrier film density and dielectric constant, and b) copper barrier film density and blanket etch rate

In contrast, our approach of adjusting process chemistry by additional C-doping reduces k and the dry etch rate while maintaining the integrity of barrier film density. This is very desirable since the barrier thickness does not have to be increased when k goes below 4.0.

Oxidation Resistance

To successfully integrate a carbon-rich barrier film, film stability and oxidation resistance are required to protect the Cu surface during various subsequent integration steps. While the additional carbon rendering the low k film more hydrophobic, it may also make the film more susceptible to oxygen attack. A thermodynamic simulation is conducted to understand various C bonding types and their effect on oxidation resistance. Two types of oxidation paths are explored, including an oxygen insertion mechanism and a peroxide mechanism:

$$O_2 + A\text{-}B \rightarrow A\text{-}O\text{-}B + O \quad (1)$$

$$O_2 + A\text{-}B \rightarrow \begin{matrix} O\text{-}O \\ | \quad | \\ A\text{-}B \end{matrix} \rightarrow A\text{-}O + B\text{-}O \quad (2)$$

Figure 2 shows the thermal barrier ΔE for different bonding structures in SiCN films [2]. While most of the conventional bonds, e.g. Si-N, N-H, Si-C, are thermally stable, C=C bonds present in the film can easily react with water or oxygen causing the film more susceptible to oxidation. In this case, the C=C double bonds unfold and bridge with the O-O bonds through the peroxide mechanism. Si-H content is of another concern. We have found that process chemistry and deposition methods need to be optimized in a way to minimize these bonds.

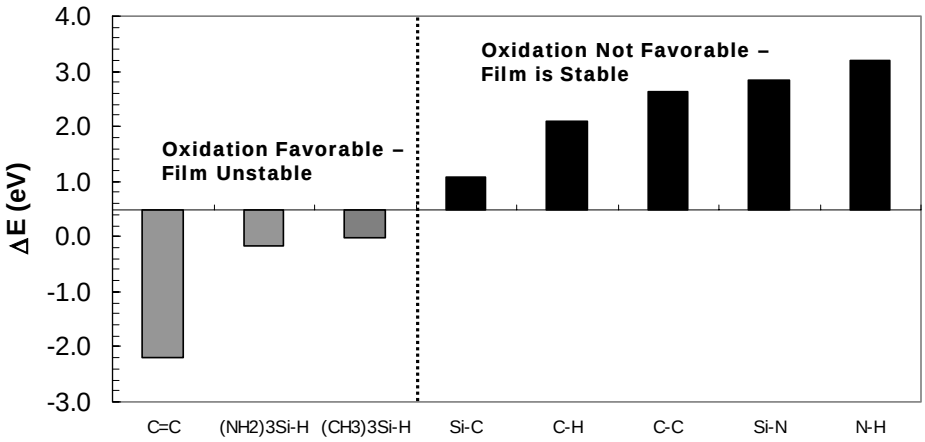


Fig. 2 Ab-initio modeling of bond oxidation and oxygen insertion into SiCN films

To verify the optimized process obtained, an aggressive test was carried out by annealing the BLOk II film capped with electroplated copper, in an ambient condition of 310°C for 24 hours. The O, C and Cu profiles were obtained using XPS analysis. In order to preserve the surface information, a SiN capping layer was added prior to XPS analysis. The results are displayed in Fig. 3 below.

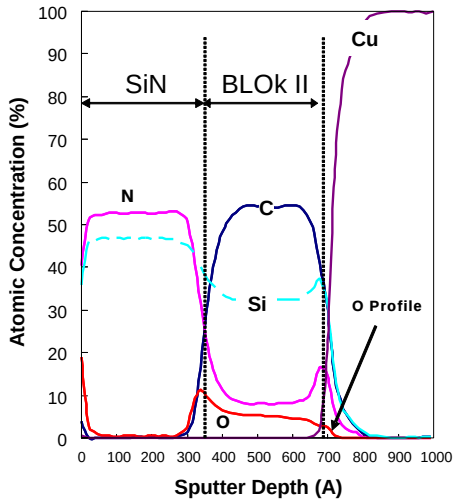


Fig. 3 XPS profile post the oxidation test described in text showing low O content at the Cu/barrier interface, comparable to that of 65nm SiCN baseline

A very low oxygen level is detected especially at the BLOk II-Cu interface. The film also contained high levels of carbon due to the process chemistry, as expected. A moisture cooker test is also conducted using an 85/85 test methodology (85% moisture at 85°C for

Ex. 1016, Page 623

17 hours) to determine the film stability no distinguishable k increase is observed. The 85/85 test is also used to verify the film hermeticity, a measure of moisture penetration through the film. Several key film properties, comparing the lower k barrier film BLOk II with that of the 65nm SiCN baseline, are listed in Table I.

TABLE I SiCN Film Properties

Film Type	**65nm SiCN Baseline**	**BLOK II**
Dielectric Constant	5.1 ± 0.1	3.8 ± 0.1
Leakage at 1MV/cm (A/m^2)	2.0 E-10	6.0E-11
Breakdown (MV/cm)	> 4.5	> 4.5
Compressive Film Stress (MPa)	250 ± 50	200 ± 50
Hardness (GPa)	10.9 ± 0.5	5.3 ± 0.5
Young's Modulus (GPa)	82 ± 3	38 ± 3
Cu Diffusion Length (nm)	< 15	< 15
Cu Oxidation Test (Ambient Anneal)	2% -O at Cu/barrier interface	2% -O at Cu/barrier interface

Mechanically, both SiCN films have much higher hardness and Young's modulus values than those of typical IMD bulk SiCO materials, and therefore can strengthen the overall dielectric stacks. In order to compensate for the tensile stress associated with the low k IMD films, the barrier dielectric can be tuned to achieve compressive stress, as shown in Table I. Except for the difference in dielectric constants, these two films possess very similar electrical and Cu barrier properties.

Interface Management

The copper surface condition prior to barrier deposition is an important factor for good reliability. It is known that Cu tends to oxidize when exposed to air. It is a common practice to add an NH_3 plasma treatment step to remove the CuO by the following reaction:

$$CuO(s) + NH_3 \xrightarrow{RF} Cu(s) + H_2O + NO_2 \quad (3)$$

In addition, this NH_3 plasma can also clean surface organic residues from the prior CMP process. Early studies have shown a clear correlation between EM activation energy and pre-treatment conditions, which impacts the SiCN adhesion to Cu, as shown in Fig. 4. The optimization of copper surface pre-treatment is obtained through plasma density adjustment and dilution of N_2. A high NH_3/N_2 ratio and high plasma density improves Cu surface cleaning efficiency. Treatment time is another knob with the tradeoff of low k damage since most IMD bulk low k materials are composed of C-doped oxides. Hydrogen ions generated from the CuO removal step can remove C from the IMD and increase its k value. This becomes more critical as the industry adopts porous SiCO film for sub-45nm node. Therefore an additional interface enhancement approach is required to mitigate the negative impact. Through XPS surface analysis of Fig. 5, we have found that the surface condition, especially the C and O content, has a large impact on adhesion strength. Figure 5 shows that Cond. III exhibits the highest adhesion strength, whereas Cond. II and I are progressively worse, indicating that lower C and O levels improve the adhesion strength. By inspecting the BLOk II XPS profile in Fig. 3, we have realized that the C presence at the interface from the deposition chemistry needs to be minimized.

Ex. 1016, Page 624

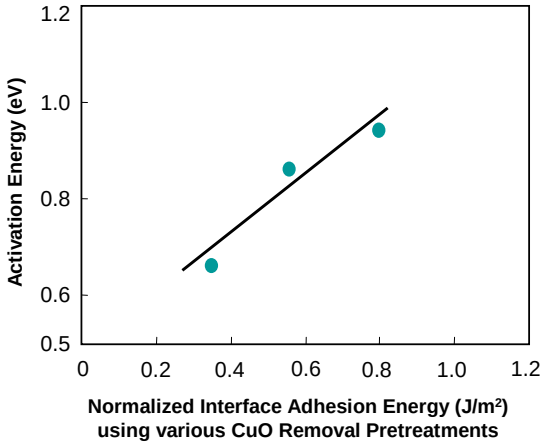


Fig. 4 Cu EM activation energy performance as a function of Cu/barrier interface adhesion strength

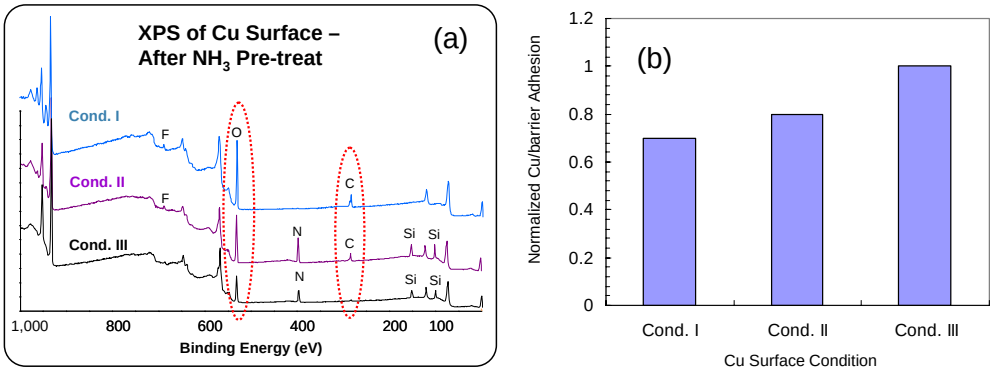


Fig. 5 Copper surface analysis and its relationship with adhesion strength to SiCN.

Adding a SiN interface sandwich layer in between, later termed Enhanced Nitride Interface (ENI), reduces C/O impurities at the interface significantly, as illustrated in Fig. 6.

In addition to adhesion improvement, ENI also provides an improved leakage and breakdown performance. ENI has been shown to improve step coverage and help to form a passivation layer to compensate the surface topography, especially at the corner area where Cu and metal barrier (Ta/TaN), IMD and SiCN converge, as shown in the TEM picture in Fig. 6. Benefits of ENI can also be verified by the electrical tests, shown in Fig. 7. The test vehicle for this study is a mask set with metal 1 CD pitch of 0.22µm, metal 2 of 0.25µm pitch and via of 0.125µm. Lithography was done using 193nm resist. The inter-layer and the intra-layer dielectric is a SiCO film with k ~ 3.0. Line-to-line I-V breakdown and TDDB tests are conducted at 105°C. Electromigration tests are carried out on packaged samples at 300ºC and a current density of 1.5mA/cm^2 using single via structures. The failure criterion is a 10% increase in resistance. Both TDDB and EM results exhibit major improvements with the SiN interface layer, as shown in Fig. 7

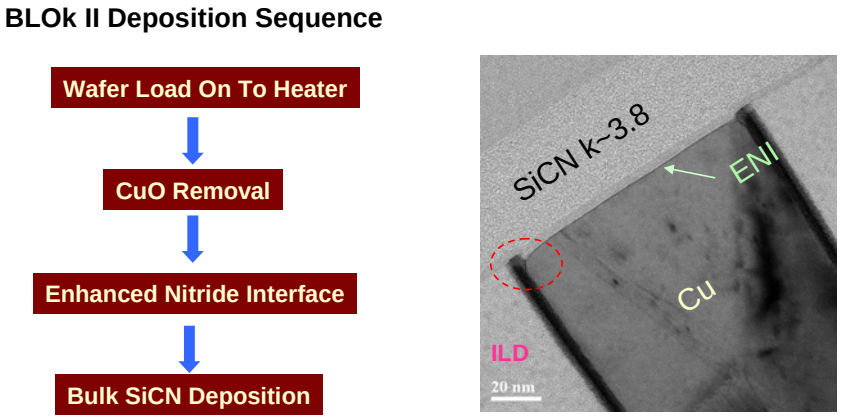


Fig. 6 BLOk II deposition sequence with ENI is shown on the left-hand side. The circle (on the right-hand-side of the graph) indicates the weak corner where 4 different materials (SiCO, Ta/TaN, Cu and dielectric barrier) meet. ENI interface covers the topographical surface prior to SiCN deposition

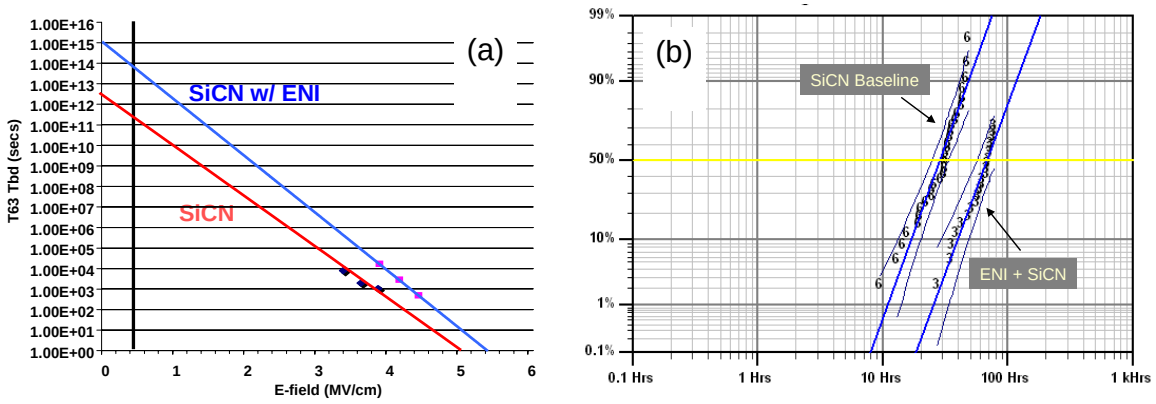


Fig. 7 Electrical tests for SiCN barrier: a) TDDB using single level Cu test structure, and b) EM improvement using ENI on 2-level Cu damascene structure

Through the ENI optimization, greater than 100% EM improvement has been demonstrated using a 2-level metal damascene structure. In addition, we have also verified that ENI has no impact on Cu line Resistivity since ENI is deposited onto the clean, pre-treated Cu surface.

Conclusions

BLOk II is an improved copper barrier for sub-45nm devices with a reduced k-value and no degradation to barrier properties when compared to the current 65nm SiCN baseline. The film structure has been optimized to minimize unstable bonding and to achieve good oxidation resistance. An Enhanced Nitride Interface process is adopted to improve the interface adhesion strength and the leakage performance. Improvements in TDDB and over 100% EM performance enhancement are also demonstrated through 2-level metal integration tests.

Ex. 1016, Page 626

References

1. S.G. Lee, Y. J. Kim, S. P. Lee, H.-S. Oh; S. J. Lee, M. Kim; I.-G. Kim; J.-H. Kim, H.-J. Shin, J.-G. Hong, H.-D. Lee, and H.-K. Kang, *Jpn. J. Appl. Phys*., 40, 2663 (2001).
2. H. Xu, M. Naik, L.-Q. Xia, V. Zubkov , R. Bhatia, C. Peterson, and H. M'Saad, *Advanced Metallization Conference Proceedings*, p. 512, San Diego, CA (2007).
3. M. W. Lane, E. G. Liniger, and J. R. Lloyd, *J. App. Phys.*, **93**, 1417, (2003)
4. A.S. Lee, A. Lakshmanan, N. Rajagopalan, Z. Cui, M. Le, L.-Q. Xia, B. H. Kim and H. M'Saad, *Mat. Res. Soc. Symp. Proc*., **812**, F5.10.1 (2004).

ECS Transactions, 18 (1) 601-604 (2009)
10.1149/1.3096508 © The Electrochemical Society

Studies of Wafer Backside Micro-arcing Prevention In Encore Ta Process

Andy Chen; Allen Bian; Paul-Chang Lin, Charles Xing; Hua Zhou*

SMIC, 18, Zhangjiang Rd, Pudong, Shanghai 201203,PRC

andy.chen@smics.com; Allen_Bian@smics.com; Charles_Xing@smics.com
JohnsonL@smics.com; Hua_Zhou@smics.com
* Correspondent author

Abstract

In advance Cu BEOL process, Encore Ta/TaN deposition process is widely used for barrier layer deposition to achieve lower via resistance and better electro migration and stress migration performance. However, wafer micro arcing existed frequently due to complex chamber design and DC/RF/AC/ESC adoption in Encore Ta process. To solve the micro arcing issue is the one of key topics in Encore Ta process. In this study, the formation mechanism of micro arcing is analyzed. Gas plug is introduced into Encore Ta process to prevent micro arcing.

Introduction

Copper metallization is utilized worldwide because of its lower resistivity and higher resistance to stress- and electro-migration, comparing with aluminum alloy. To prevent Cu diffusion into dielectric, metal layer Ta/TaN is utilized as a barrier layer. Encore Ta process, an upgrade version based on SIP (Self Induced Plasma), is adopted in 90nm and beyond technology because of better gap fill performance to meet higher aspect ratio and thus, better via resistance. [1] However, the complex chamber configuration may induce several side effects, one of the serious side effects is wafer backside micro arcing during chamber process. The micro arcing may damage device and kill chip Cp yield. [2] It is very necessary for Cu line process to analyze the arcing formation mechanism and solve this issue. In this paper, the formation mechanism of wafer backside micro arcing is analyzed. Studies show the formation of wafer backside micro arcing mainly related with wafer backside plasma condition, and the defect hit-rate can be significantly reduced through gas plug installation.

Experiment

Encore Ta process is utilized to deposit Ta/TaN barrier layers in AMAT Endura 5500 tools .A supper low temperature Electrostatic Chuck (E-chuck) is adopted for wafer temperature control. The wafer is located on the E-chuck (See Fig.1) while the deposition process begins. For Encore Ta process, high radiofrequency power is used to tune the plasma re-sputtering, which can optimize film step coverage and uniformity. [3]

Supper low temperature E-chuck is a kind of bi-polar chuck, in which the contact surface is deposited one thin dielectric layer. To avoid processed wafer from over temperature, IHC (insulated Helium cooling) system is adopted to keep wafer temperature under control. Wafer backside Argon gas, which is used to keep the

Ex. 1016, Page 629

processed wafer temperature low and uniform, goes through the E-chuck and actives on the wafer backside. During process, DC/RF/AC/ESC powers are all on to assure wafer under uniform plasma circumstance.

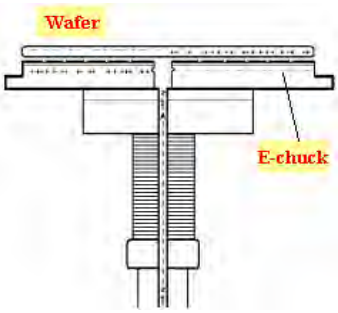


Fig 1 wafer located on E-chuck

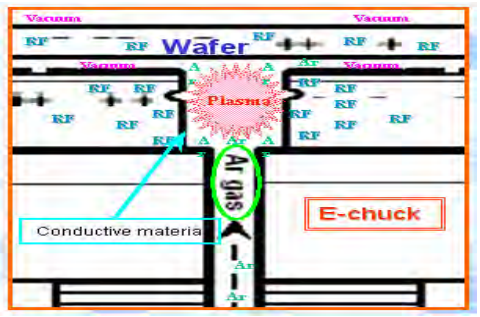


Fig 2 Mini Plasma condition formation

In this experiment, the wafer backside arcing material is used to study through OM (optical microscope), and FIB (Focus Ion Beam). The new idea of gas plug will be tested, 8 inch-P- type wafers is used for Ta/TaN deposition process.

Results and Discussion

One of mini plasma condition is forming at the backside gas outlet. (See Fig 2) According to PVD theory, the mini plasma will be deposited at the backside of processed wafer, which might lead to wafer backside arcing.

The map of wafer backside center arcing is showed in Fig 3, It looks like that there is some discolored material with arcing mark sticking on the backside of wafer center. And FIB shows the material layer is Ti element in Fig4.It is obvious that arcing happened at the wafer backside center and an extra layer of Ti was deposited while Encore Ta/TaN process is on. Wafer backside arcing will impact the chip structure and wafer Cp yield loss; Furthermore, wafer may be broken at the subsequence Cu ECP (Electro chemical plating) anneal process.



Fig 3.Wafer backside arcing OM review

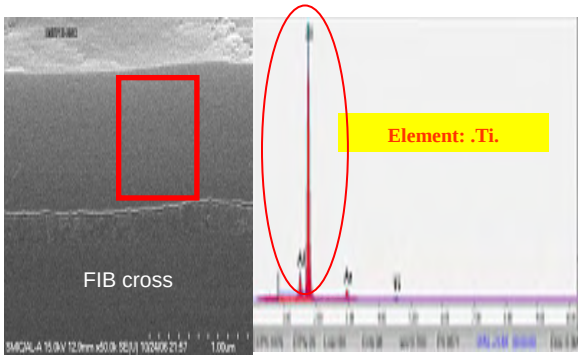


Fig 4 Wafer backside arcing material FIB review

E-chuck surface is deposited with dielectric layer and dimple, and the dimple material is mainly composed of Ti.Due to the gap of E-chuck gas outlet is too big, there is more residue gas in this area at the processed chamber; Under the complex and strong AC /RF bias power application on wafer, it will form a mini-plasma condition, causing a Ti metal film deposited at the wafer back side. When DC/AC / RF power simultaneously turns on, high voltage difference between wafer backside Ti metal film and E-chuck center will induce arcing through the weakest point.

Wafer backside Ti film deposition is usually found after Encore Ta/TaN process that suffers wafer backside arcing issue. A detailed investigation can reveal that E-chuck gas outlet center surface showed the seriously arcing mark in Fig 5.



Fig 5 E-chuck top view structure

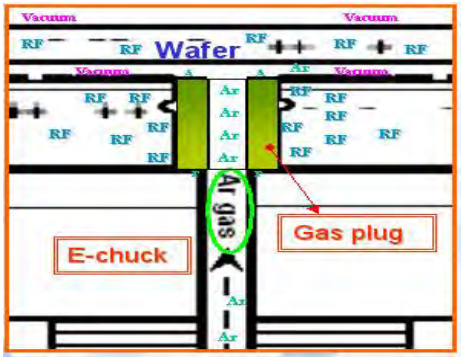


Fig 6 gas plug installation picture

From analysis above, mini-plasma condition formed at E-chuck gas outlet center is the root cause for wafer backside arcing and Ti deposition. In order to eliminate this mini-plasma condition and thus prevent processed wafer backside micro arcing, an innovative Gas plug made of insulated material is developed and inserted into the E-chuck gas outlet center in Fig 6. Compared Fig 2 with Fig 6,there is no gap for residue gas in Fig6. At this circumstance, there is no mini-plasma condition at the gas outlet center, and throughly prevent wafer backside micro arcing with E-chuck surface.

Conclusions

Gas plug is a very innovative idea, which can scale down the E-chuck gas outlet center gap, and prevent the mini-plasma condition forming. So there is no deposition film at processed wafer backside and no any arcing happened with the E-chuck surface. After mass application on the production line, it proves that it is very practical and effective. No other side effect has been observed.

Acknowledgements

Authors would like to thank our supervisors Allan Zhou, Division head of SMIC-S1-TF and Daniel_Fu, Center head of SMIC-S1 center, for their encouragement and guidance.

References

[1] H, Kin; A.J.Kellock; S.M.Rossnagel, J.Appl.Phys.98, 01438(2005)
[2] Kun, Li; Er, E; Yeow, T; Dong Tang, Device and Material Reliability. IEEE Transaction on Volume 6,Issue 2,June 2006,Pages (283-287)
[3] G. Bersuker, V. Blaschke, S. Choi and D. Wick, conduction Mechanisms in Cu/Low-K Interconnect, 0-7803-5649-7199, *IEEE*, 1999.

Ex. 1016, Page 632

ECS Transactions, 18 (1) 605-610 (2009)
10.1149/1.3096509 © The Electrochemical Society

Pre-clean Study for Both Damascene SiN and N-Blok to Resolve Queue-time-dependent Dielectric Barrier VBD Degradation

Jerry Xu[1], Rong Gao[1], Hongjie Wang[1], Xianyuan Li[1],
Sang Lee[2], David Cui[2], Mehul Naik[2], Li-qun Xia[2],
Qiang Xu[3], Ziying Zhang[3], and Emily Bei[3]

[1]Applied Materials, Shanghai, China
[2]Applied Materials, Santa Clara, USA
[3]Semiconductor Manufacturing International Corporation, Beijing, China

Abstract

Recently, it was found that for 65 nm logic device the VBD for SiN decreased dramatically when the queue time after Cu CMP (prior to damascene SiN deposition) is longer than two hours. In addition, this degradation increases with queue time. The root cause has been isolated to the interface between SiN and BD I interface at M5. Detailed analysis revealed that the change of Cu oxidation post Cu CMP is responsible for the degradation. We therefore have developed a series of new pre-clean recipes to tackle the issue. In this paper, we will discuss the potential mechanisms of time dependant VBD degradation and the practical solution to this problem.

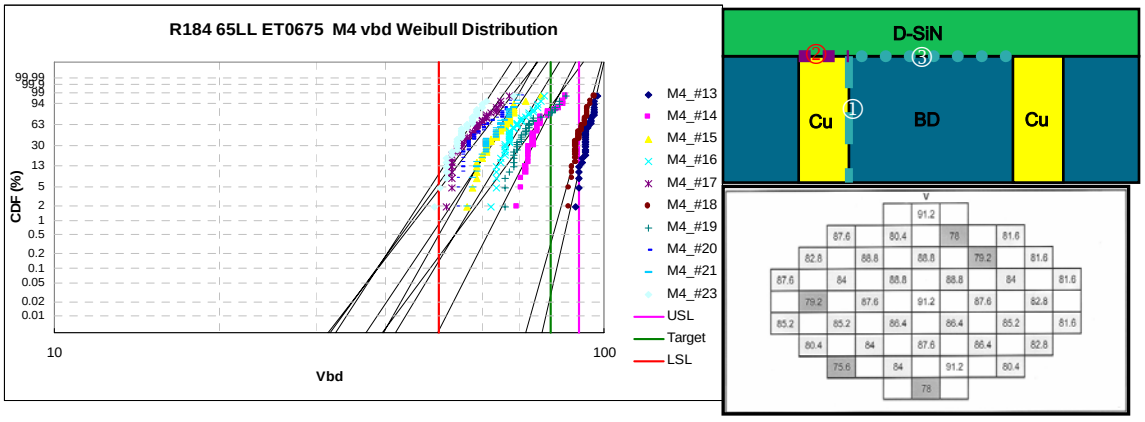
Introduction

As semiconductor logic device scaled down to 65 nm, new materials and technologies are introduced to meet the stringent scaling requirement. For example, NiSix as a new material has to be introduced as salicide to replace the existent CoSix. Also, stress engineering is implemented to overcome the NMOS mobility degradation problem. Many corresponding challenges start to surface, as expected. For a typical dual damascene back end of line processes, certain unforeseen challenges also show up. For 65 nm logic devices, low k dielectrics like Black Diamond I (BD I, k=3.0) and barrier for low k (Blok I, k=5.0) are used for lower layers of metal, e.g., from M1 to M5. For higher metal layers, fluorine doped silicon oxide (FSG) and SiN are applied. However, at M5, there is a new, never encountered interface between BD I and SiN. Recently, it was found that the break down voltage VBD for this metal layer started to degrade when the queue time (post M5 Cu CMP and before SiN barrier deposition) was longer than two hours. In addition, the degradation increases with increasing the queue time, e.g., VBD decreased from 60 V to less than 30 V when queue time was longer than 24 hours. In this paper, this problem is discussed and the corresponding solution is proposed as well.

Ex. 1016, Page 633

Problem Statement

For SMIC's 65 nm low leakage product, there is one major queue-time-dependent voltage breakdown of dielectric (VBD) degradation problem for both damascene and N-Blok. When the queue time after Cu CMP (prior to damascene SiN or N-Blok deposition) exceeds two hours, the VBD for SiN decreases dramatically (from about 60 V to below 30-40 V if the queue is longer than eight hours), as shown in Fig. 1. It is worth noting that the only wafers that passed the specification of 50 V were the ones with one hour queue time. The problem was suspected to be originated at the SiN and BD I interface (Metal layer 5 with the same design rule of line spacing of Metal layer 2). The resulting issue impacted BEOL production significantly.



(a) (b)

Fig. 1 (a) The VBD values for all the dies on the wafer were plotted with a Weibull distribution. All the dots for one wafer were fitted with a linear straight line. The interception of this line with 0.1% population line (parallel to x-axis) is defined as the VBD value for the wafer. The specification for the VBD is 50 V (the red vertical line). As shown above, only two wafers or lines (blue and brown) met specification, (b) film structure and die map distribution. Most of the failed dies are found to be close to wafer edges.

Results and Discussions

I. Hardware check and bulk film property adjustment

Low VBD values tend to cause EM (electromigration) issues. Various models were introduced to explain VBD degradation, including hardware drift and fab environment. We performed thorough hardware audit, changed new face plates and blockers, installed filters (acid/base/organic/particle) at FI, verified all the MFCs, RF power, resistivity of CVG, pumping curve, SiH_4 MFC delay sequence (to avoid SiH_4 burst for CuSix formation on top of the Cu trenches, shortening the dielectric distance between two adjacent trenches and lead to lower VBD), etc., to find no smoking gun. We performed the same splits on a 2nd system and observed the same results. We later optimized the bulk film properties (e.g. to reduce the SiN film's reflective index (RI)) and implemented

Ex. 1016, Page 634

both bilayer SiN (with a lower deposition rate SiN in the beginning) and low RI SiN. The VBD degradation improved only slightly with low RI SiN better than bilayer and baseline, as shown in Fig. 2. At this stage, it was concluded that bulk film properties and hardware drift were not the main factors for VBD degradation. The efforts were later shifted to interfacial quality between SiN/BD I or between SiN/Cu. Hydrogen-contained plasma treatments are the key knobs chosen to see if VBD can be modulated from modifying the interfacial properties.

No.	Tool ID	Split condition	Q-time	#13	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23
Test 2	SiN02	SIN750-20ANL(BL)	1	82					78					
		SIN750-20ANL(BL)	4		57					48				
		SIN750-20ANL(BL)	8			44					39			
		TLRSIN750 (Low RI SiN)	8				49					46		
		SIN750-20ANL(BL)	12					39						39

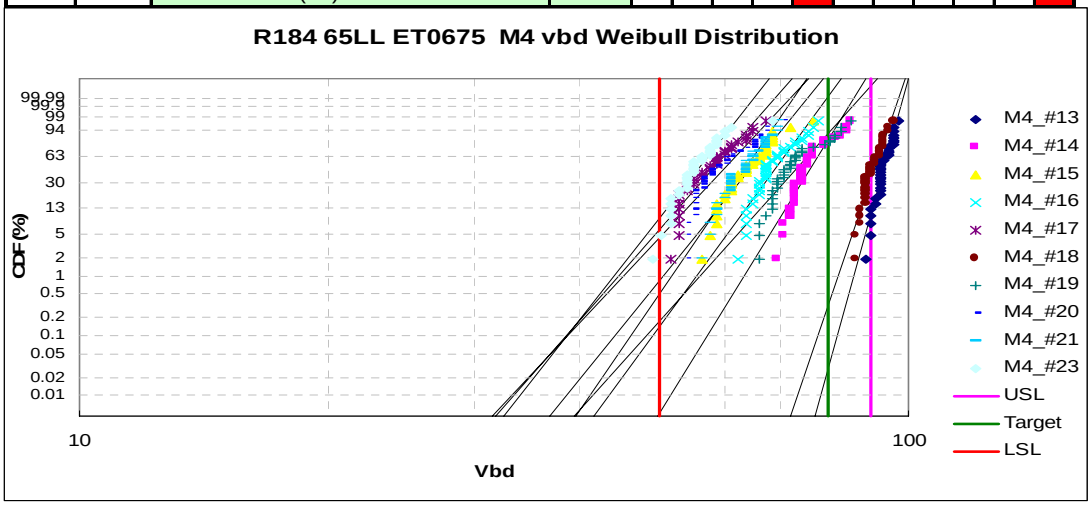


Fig. 2 VBD did not meet 8 hours queue time specification ($t_{0.1}$>50V) even in newly audited chamber and a 2nd system, which indicated the VBD issue is not hardware related; Low RI SiN improved VBD, but not enough to meet specification and bilayer SiN gave worse result.

PECVD based pre-clean was implemented to remove CuOx prior to the deposition of dielectric barrier since the introduction of Cu/Oxide at back-end-of-line for logic devices in 1990's. Various gases and combinations have been developed, such as diluted H_2 gas, diluted NH_3 gas, pure NH_3 gas, pure NH_3, and pure He gases. H_2 based gas and pure He plasma treatments were obsolete due to the strong plasma induced damage, which caused severe gate oxide integrity issues. So far, only NH_3 based plasma clean is used for both damascene SiN and N-Blok. A diluted NH_3 plasma with low power is implemented for SiN and N-Blok in order to avoid the potential problem of severe Cu re-sputter (onto the faceplate). Such an approach was proven to be effective till 90/65 nm logic devices. However, this diluted low-power NH_3 plasma could be too weak for the current process.

II. Additional root causes and the corresponding solutions

Additional potential root causes for the VBD degradation discussed here are identified, which include (1) pollution absorption on BD I surface during the queue time, (2) CuOx degradation during queue time, (3) low efficiency of CuOx removal or even Cu sputter for BKM pre-clean conditions, (4) the plasma damage on the BD I surface which leads to possible leakage pass ways, (5) the usage of EC clean at the end of Cu CMP, and the potential remaining of acid on Cu surface could lead to corrosion during queue time. Extensive splits on patterned wafers have been performed at customer sites and at the 300 mm fab of Applied Materials in US to isolate the culprit.

The results from Applied Materials' lab confirmed for the first time that VBD degradation is dependent strongly on the queue-time, as shown in Fig. 3. In addition, the VBD value recovered back to original number (the red triangle in the plot below) with Cu CMP touch up on the wafer surface (after queue time). This proved that the bulk film property should have nothing to do with this problem. Further test with wafers placed in N2 box showed almost no VBD degradation during queue time (Fig. 4). This suggested that fab environment or pollution was not the root cause. Extensive Cu surface oxidation or the copper oxides property degradation with time could be the real root cause.

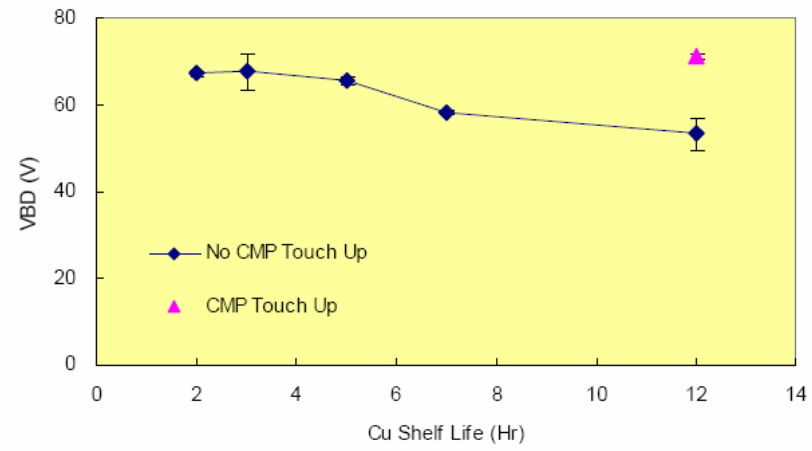


Fig. 3 Queue time has impact on VBD at closer spacing; CMP touch up recovered VBD, which indicate Cu surface change overtime may be the root cause for VBD degradation (Test Structure: 45nm oxide/SiN structure; Oxide instead of BD I is on purpose to exclude potential BD I damage impact on VBD).

Environment	Shelf-life	75nm Line Spacing		85nm Line Spacing	
		1st (V)	6th (V)	1st (V)	6th (V)
Air	3 days in air	26.8	28.8	39	40.7
Air	6 days in air	25.5	25.6	37.3	37.2
Dry N2	CMP touch-up after 6 days in air	50.4	47.9	65.7	63.3
Dry N2	10Hr after CMP touch-up	47.8	47.9	60.9	63.8
Dry N2	6 days after CMP touch-up	46.4	46.1	64	63.8

* Bi-layer BLOk deposition, D-NH_3 treatment

Ex. 1016, Page 636

Fig. 4 Put lot in N2 box did not see obvious VBD degradation, which indicates Cu oxidation in air maybe the root cause for VBD degradation.

However, it was interesting to note that both higher power and lower pressure clean led to less BD I damage (as measured from k value shift), as shown in Fig. 5, and therefore less VBD degradation. This has been confirmed with the results - higher RF power and lower pressure does result in less VBD degradation and therefore higher CuOx removal efficiency, as shown in Fig. 6. Also, it was found that Cu touch up (CMP) at the end of 8 hour queue time could recover the high VBD value (Fig. 3), which validated that Cu surface change over time is the root cause.

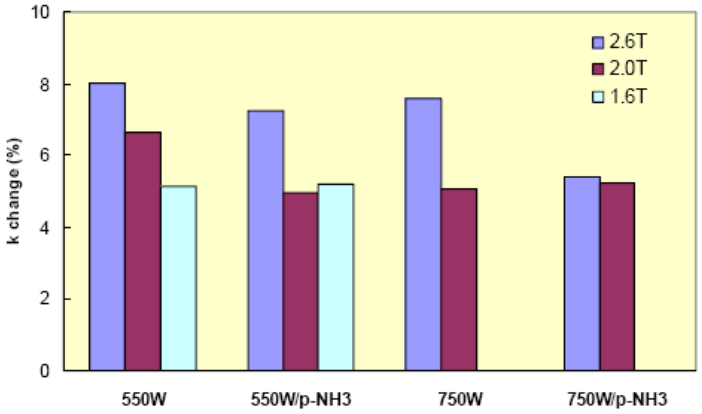


Fig. 5 High power combined with pure NH3 clean gives the lowest k shift and low pressure helps at lower RF power.

No.	Tool ID	Split condition	Q-time	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23
Test 1	SIN02	TSIN750-20ANL (Baseline)	?	54																						
		TSIN750-5PT (5s treatment)	?		51																					
		TSIN750-40PT (40s treatment)	?			59					62															
		TSIN750-550RF (550w Power)	?				31					40														
		TSIN750-PURE (pure NH3 treatment)	?					56					58													
		TSIN750-5STAB (5s stable)	?						53					50												
		TSIN750-2PD (2s pre-dep)	?							57					62											
Test 2	SiN02	SIN750-20ANL(BL)	1													82					78					
		SIN750-20ANL(BL)	4														57					48				
		SIN750-20ANL(BL)	8															44					39			
		TLRSIN750 (Low RI SiN)	8																49					46		
		SIN750-20ANL(BL)	12																	39						39
Test 3	SiN02	TSIN750-20ANL (Baseline)	1	65							63															
		TSIN750-20ANL (Baseline)	8		48							51														
	SIN01	TSIN750-20ANL (Baseline)	8			48							40					V					V			
		4.2T,550W,990sccm,0sccm	8				47							V					V					V		
		3.7T,550W,1000sccm,10000sccm	8					34							V					V					20	
		2.6T,550W,1700sccm,700sccm	8						48							V					V					51
Test 4	SiN01	TSIN750-20ANL (Baseline)	1	65																						
		TSIN750-20ANL (Baseline)	8		30						40															
		Pure NH3 treatment + Low RI SiN	8			46				44																
		2.6 torr NH3 treatment + Low RI SiN	8				52		48																	
		Pure NH3 treatment + Bi-layer SiN	8					42				44														

Fig. 6 Bi-layer did not see VBD improvement in S-Company. Best results were achieved by using 2.6T low pressure treatment and low RI SiN combination. However, the net benefit is still marginal.

An optimized recipe, with 2.6 torr, 550 W and 1700 sccm of NH_3 with 700 sccm N_2, showed the best results – at 8 hour queue time, one wafer had $t_{0.1}$=52 V and the other

$t_{0.1}$=48 V, as shown in Fig. 6. This recipe was later implemented for N-Blok and satisfactory results of $t_{0.1}$=58 V and the other one of $t_{0.1}$=56 V at queue time of 8 hours were achieved, as shown in Fig. 7. New splits are under the way to confirm the new pre-clean recipe results and to use even lower pressure, higher NH_3 partial pressure, low frequency RF (to verify root cause number (4) because low frequency RF will induce higher plasma bombardment on BD I surface) and the use of basic solution for post Cu CMP clean (to verify root cause number (5)).

No.	Tool ID	Split condition	Q-time	Pair	Result	
Test 2	GTLOK04 CHA	T-NDC450-A (Baseline: Pure NH3 treatment)	8	6th	53	52
		N2 dilute NH3 treatment + LR BLOK	8	1st	43	
		N2 dilute NH3 treatment + LR BLOK	8	6th	42	44
		2.6T/550W/1700sccm/700sccm + LR BLOK	8	6th	58	56
		Bi-layer (thin SiN+LR BLOK)	8	6th	46	47

Fig. 7 2.6 torr/550W pre-clean recipe together with low RI Blok showed the best VBD results at queue time equal to 8 hours.

Conclusions

In this paper, a new phenomenon of queue-time dependent dielectric VBD degradation was observed for a Cu 65 nm logic device BEOL process. The corresponding root causes were analyzed to find that the degradation of CuOx is the culprit. New pre-clean solutions were then developed to improve the efficiency of CuOx removal. These results could also be useful for other logic devices beyond 65 nm.

Acknowledgments

We would like to thanks Mr. Ming Cai, Mr. Yalei Hu and Dr. Jay Ning at Semiconductor Manufacturing International Corporation in Beijing and also Mr. Yingjie Chen, Mr. James Chen at Applied Materials China Account for their assistances, valuable suggestions and discussions.

ECS Transactions, 18 (1) 611-615 (2009)
10.1149/1.3096510 © The Electrochemical Society

A Study on PVD Co Pre-clean Chamber RF Reflect Power Improvement

Kun Gui, Pin-Sheng Hu, Dong Ouyang, Chi-Hsien Peng,
Huang-Her Yang, and Paul-Chang Lin

SMIC (Semiconductor Manufacturing International Corp)
18, Zhangjiang Rd, Pudong, Shanghai, PRC, 201203

Abstract

For 0.18um and 0.13 um technology, Co/TiN deposition process is widely used to form CoSix to provide the needed lower contact resistance of the polysilicon gate. CoSix formation is sensitive to pre-existing wafer surface native oxides and ambient contamination. Pre-clean process before Co film deposition is required to remove residues before Co/TiN deposition. The corresponding chamber RF reflected power too high issue during this pre-clean has significant impact on the etch rate and particle performance. In this paper, three approaches to reduce Co pre-clean chamber RF reflect power are presented.

Introduction

The PCII chamber uses an RF resonator and two RF generators of different frequencies. It applies a non-reactive Ar plasma to etch the wafer lightly before the Co/TiN deposition. In general, a thin layer (around 30 ~ 60A) of oxide is expected to be removed by this Ar plasma bombardment. It is well known that high energy Ar plasma may sometimes induce high leakage current and damage the device [1]. It is therefore paramount to maintain etching performance by keeping the RF power stable.

For the thinner Co/TiN production, it is a significant challenge to maintain the RF power stable under the requirement of very short etch time (~ several seconds), low RF power, and low RF power ignition pressure. Fig. 1 shows an example which the RF reflected power is too high at low pressure and the process is forced to abort. An experiment is then set up to tackle the problem, as detailed in the following sessions.

Ex. 1016, Page 639

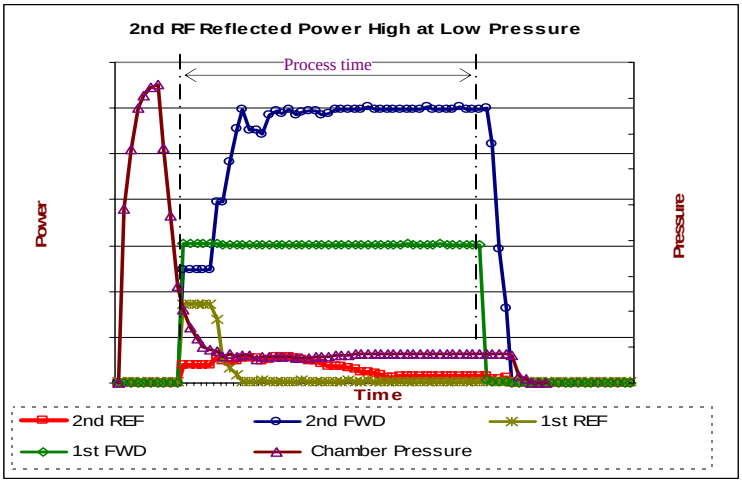


Fig.1 An example of reflected power too high

RF reflected power is generated by the difference between the impedance of the power cable (Zo=50 Ohm) and the chamber impedance (Zr). The RF match network is designed to adjust the impedance of chamber by adjusting the matching capacitor and the inductor. If the chamber impedance cannot match Zo, an RF reflected power will be generated. If this RF reflected power is too high, the plasma will not be stable enough and the system will have to abort. This implies that the impedance matching of the chamber is critical for RF performance. There are a few well-known factors impacting the impedance of the chamber, such as gases, temperature, and even clearness. In addition, the RF generator and RF matchbox can also strongly influence the RF reflected power. To reduce Co pre-clean chamber RF reflect power, we have proposed three approaches: 1) add an "Ignition" step as pre plasma before main etch, 2) optimize two stage RF power ignition point, and 3) optimize the spacing.

Experiment

The recipe optimization experiments mentioned above were carried out on an AMAT Endura 5500 system. Fig. 2(a) shows a schematic of the ICP Pre-clean chamber. The 1st RF power uses Hi-Frequency RF Power to form a DC bias and to enhance the etch rate. The 2nd RF power uses a Low Frequency RF Power to create inductive-coupled plasma. Fig. 2(b) shows the equivalent circuit diagram.

Ex. 1016, Page 640

(a) (b)

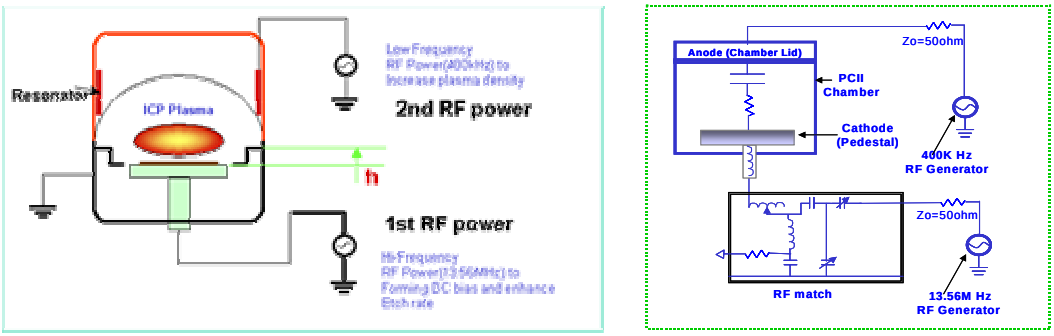


Fig. 2 Sketch map of ICP Pre-clean chamber and equivalent Circuit diagram. [2]

Result and Discussion

Chamber pressure at plasma ignition impacts chamber impedance significantly. CoSiX pre clean RF power ignites at lower pressures. This lower pressure condition increases the plasma impedances and renders it difficult to ignite the plasma. RF plasma ignites easily at higher pressures. We therefore added an RF power "Ignition" step, in which the chamber pressure was set at some higher value, to help igniting the plasma and minimizing the reflected power. Fig. 3 shows that the 2nd RF reflected power decreases evidently after adding this pre plasma step.

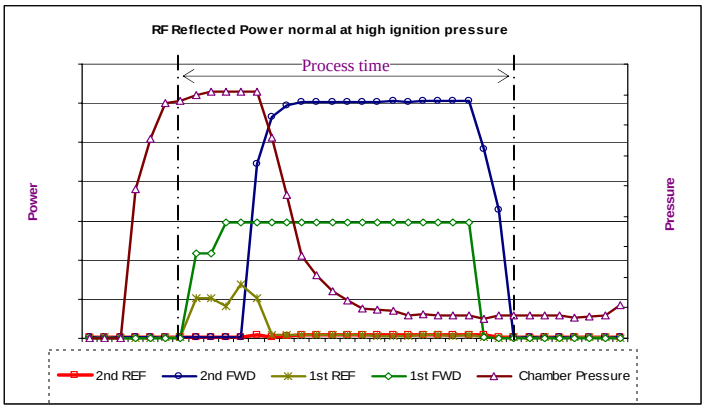
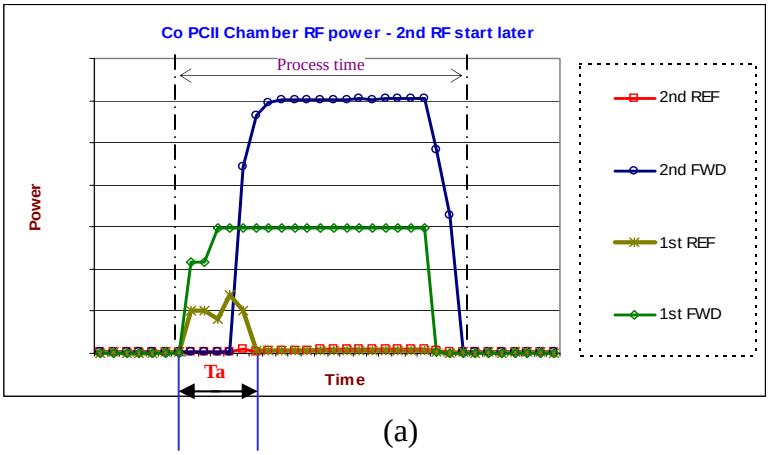


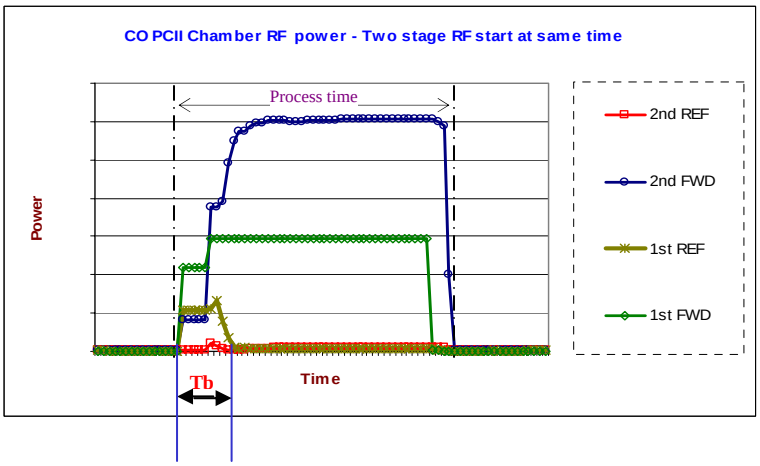
Fig.3 RF forward and reflective power profile with the new "Ignition step" and the 2nd reflected power is negligibly small.

We also found that most RF reflected power high cases happened at the beginning of the process, as shown in Fig 4(a). It is obvious that the 1st RF power ignites ahead of the 2nd RF power, and the 1st RF reflected power picks up when the 2nd RF power ignites. This

discrepancy is corrected by adjusting the two stage RF power to be turned on at the same time. Results shown in Fig. 4(b) demonstrate that 1st RF reflected power has decreased 24 % if the two-stage RF power is ignited at the same time.



(a)



(b)

Fig. 4 real time power chart of Pre clean chamber process

Chamber spacing is the 3rd knob for the optimization and is usually ignored. RF reflected power becomes more and more sensitive to chamber spacing at low ignition pressure. An optimized spacing is a function of chamber conditions and recipe parameter settings. A spacing matrix chart, shown in Fig. 5, is used to identify this optimum spacing. The uniformity, etch rate and the particle performance must be maintained during spacing fine tune. An optimized spacing of a3 is resulted with a minimum reflected power, as shown in the graph below.

Ex. 1016, Page 642

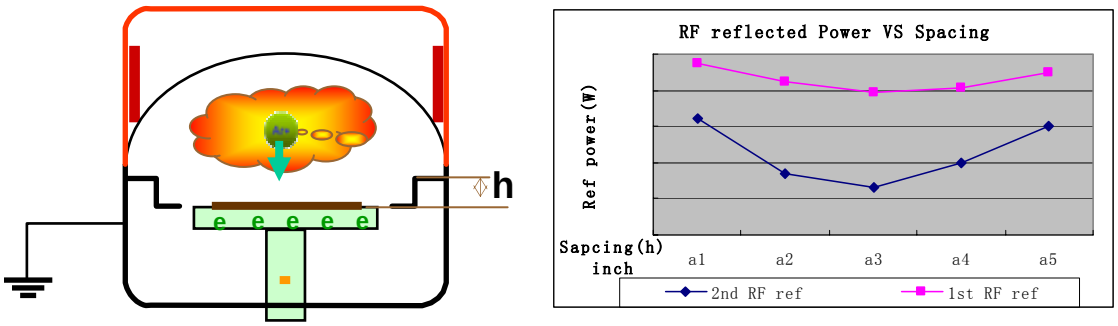


Fig.5 Chamber spacing, marked as h in the graph, effect on RF reflected power

Conclusions

This paper introduces three approaches to reduce the reflected power. Modification of two stage RF power ignition point can reduce the 1st RF reflected power effectively. Adding an "Ignition" step can improve RF power ignition and resonating. Adjusting the chamber spacing can help to optimize the two stages RF reflected power.

Acknowledgements

Authors would like to thank our supervisors Allan Zhou, Division head of SMIC-S1-TF and Daniel Fu, Center head of SMIC-S1 center, for their encouragement and guidance.

References

1. M. -H. Chi, S. Hao, C.H. Shen, H.J. Lin, C.C. Chen, Y.F. Lin, C.S. Lee, and K.H. Chang, "Leakage Current and Defect Monitors for Ultra-shallow Co-silicide Junctions in Advanced 0.13um CMOS," Semiconductor Manufacturing, 2003, IEEE 2003, p.366-369.
2. Bing Wu, "Endura HP PVD Function Description Training," Version A1, p.3-30, March 2000.

Ex. 1016, Page 643

ECS Transactions, 18 (1) 617-623 (2009)
10.1149/1.3096511 © The Electrochemical Society

Robust $Hf_xSi_{1-x}O_2$ Gate Dielectric using Co-injection Atomic Layer Deposition

Larry Bartholomew[a], Andy Bavin[a], Thierry Lazerand[a] and Vivek Rao[a]

[a] Aviza Technology, 440 Kings Village Rd., Scotts Valley, California 95066, USA

Using a novel Atomic Layer Deposition (ALD) process in which multiple chemical precursors are simultaneously injected to form a monolayer on the substrate and then oxidized, we demonstrate a robust hafnium silicate film that can be easily tuned to meet specific device requirements. A large process window is demonstrated via physical and electrical performance of the film. The film has been fully integrated into a CMOS structure and repeatable parametric results over a wide range of film compositions are shown.

Introduction

Hafnium silicate ($Hf_xSi_{1-x}O_2$) has been adopted as an alternative gate dielectric to silicon oxide (SiO_2) or silicon oxy-nitride (SiON) in advanced CMOS devices due to its high dielectric constant, band alignment to silicon, thermal stability and relative ease of integration (1). In addition, the Hf composition of the film can be varied to change the physical and electrical properties of the material making it possible to tune the device performance to meet specific requirements (2, 3).

The techniques chosen to deposit the $Hf_xSi_{1-x}O_2$ film must be capable of achieving physical film thickness in the range of 15 to 20 Å with excellent uniformity and repeatability. In addition a high purity, high density film is needed to ensure the required electrical performance. For some advanced integration schemes, there is also a step coverage requirement (4). Typically either a chemical vapor deposition (CVD) or atomic layer deposition (ALD) process is employed to deposit the high-κ dielectric. While CVD processes have a higher deposition rate than ALD, they have also been shown to produce films with lower density and therefore higher electrical leakage. The CVD reaction also lends itself to some process instability that can have an effect on the repeatability of the physical and electrical properties of the $Hf_xSi_{1-x}O_2$ gate dielectric. Small variations in the film thickness, composition or density can significantly increase the development time needed to resolve the various integration challenges due to adoption of a high-κ dielectric, and can result in a narrow process window for volume device manufacture. The ALD process uses discrete precursor and reactant pulses separated by a nonreactive gas purge to deposit a single atomic layer per process cycle. As the reaction is self limiting, ALD enables excellent thickness control and process repeatability.

ALD of binary component films such as $Hf_xSi_{1-x}O_2$ are typically deposited by sequential depositions of hafnium oxide (HfO_2) and SiO_2. This "nano-laminate" technique requires a high temperature anneal to be performed after deposition to form a high quality $Hf_xSi_{1-x}O_2$ film. As a consequence, the technique has a high thermal budget and requires an additional process step, yielding a lower overall throughput.

Through careful selection of compatible precursors it is possible to deposit $Hf_xSi_{1-x}O_2$ using a "co-injection" ALD technique. In this process, Hf and Si precursors are pulsed at the same time followed by single oxidation pulse. Co-injection halves the number of process steps required and eliminates the need for any change in process conditions

between HfO_2 and SiO_2 deposition layers while producing dense $Hf_xSi_{1-x}O_2$ films that are resistant to crystallization without the need for high temperature annealing.

In this paper, we demonstrate a robust $Hf_xSi_{1-x}O_2$ film deposited by a co-injection ALD process. The process enables repeatable deposition of $Hf_xSi_{1-x}O_2$ gate dielectrics with a wide process window.

Experimental

In order to evaluate the process stability of the $Hf_xSi_{1-x}O_2$ film a process DOE was performed in a Celsior *f*xP™ ALD tool from Aviza Technology. The system employs a showerhead and radial pumping gallery to ensure uniform precursor and reactant delivery over a 300 mm Si wafer. A schematic diagram of the chamber and chemical delivery configuration is shown in Figure 1. The reaction volume is minimized to ensure rapid pulse and purge and is heated to 150 °C. $Hf_xSi_{1-x}O_2$ was deposited using co-injected liquid Tetrakis-Di-Methyl-Amino-Hafnium (TDMAHf) and Tris-Di-Methyl-Amino-Silane (3DMASi) precursors of >99.999 % purity. The precursors were chosen for their low delivery temperature, chlorine-free chemistry and compatibility at process temperatures around 400 °C. A commercially available ozone delivery system was used to generate the ozone reactant with a concentration of 240 g/m^3 in oxygen. Argon (Ar) was used as the precursor carrier gas and purge gas.

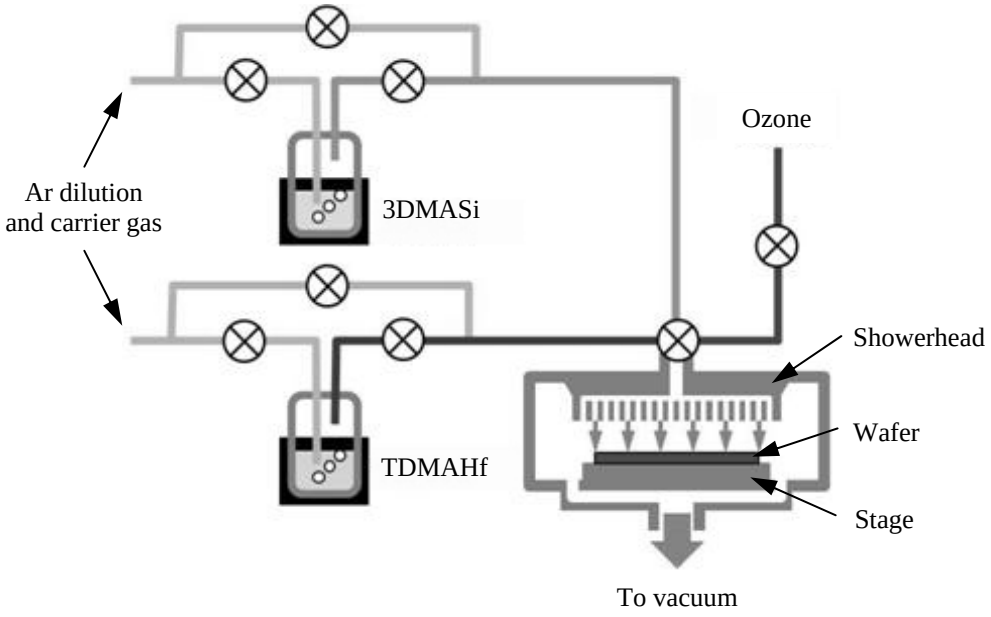


Figure 1. A schematic diagram of Celsior *f*xP™ ALD process chamber and chemical delivery system.

A central composite design 3-parameter DOE was chosen to evaluate the process window of a $Hf_{0.65}Si_{0.35}O_2$ film. A physical film thickness of 60 Å was selected to allow CET measurement on the Mercury Probe and minimize dilution effects from the substrate native oxide layer. Stage temperature was varied from 412 to 462 °C, precursor pulse time from 0.4 to 0.6 s and chamber pressure from 400 to 600 mTorr. All other process conditions remained unchanged from the baseline settings. Depositions were performed in random order on reclaimed 300 mm Si <100> wafers measured to have 12Å native

Ex. 1016, Page 646

oxide. In addition, 6 center-point conditions were processed using the baseline recipe to monitor process repeatability within the DOE.

Characterization

The physical film thickness and composition control of the $Hf_{0.65}Si_{0.35}O_2$ films were measured at 49 points with an edge exclusion of 3 mm on a Metrosol Vacuum Ultraviolet VUV4000 system (5).

Optical film thickness and refractive index were measured at 49 points with an edge exclusion of 3 mm on a KLA-Tencor F5x. Measurements of leakage current density and CET were performed on the as-deposited $Hf_{0.65}Si_{0.35}O_2$ films at 5 points with a contact area of 7.15×10^{-5} cm^2 using a Four Dimensions 3092-A Mercury Probe. For all measurements within-wafer and wafer-to-wafer uniformity was defined as the standard deviation divided by the mean value.

Results and Discussion

The results from all tests were examined to evaluate the stability of the ALD $Hf_{0.65}Si_{0.35}O_2$ process across the entire DOE matrix as shown in Figure 2. The mean physical film thickness from all tests was 62.2 Å with all wafers under 1.71 % within-wafer uniformity and a wafer-to-wafer uniformity of 3.57 %. The mean Hf composition was 64.8 % with all wafers under 2.98 % within-wafer uniformity and a wafer-to-wafer uniformity of 1.50 %. The mean refractive index across all wafers was 1.684 with <0.64 % within-wafer uniformity and a wafer-to-wafer uniformity of 0.21 %; the leakage current ranged from 9.5×10^{-9} to 1.4×10^{-8} A/cm^2 measured at -1.5 V.

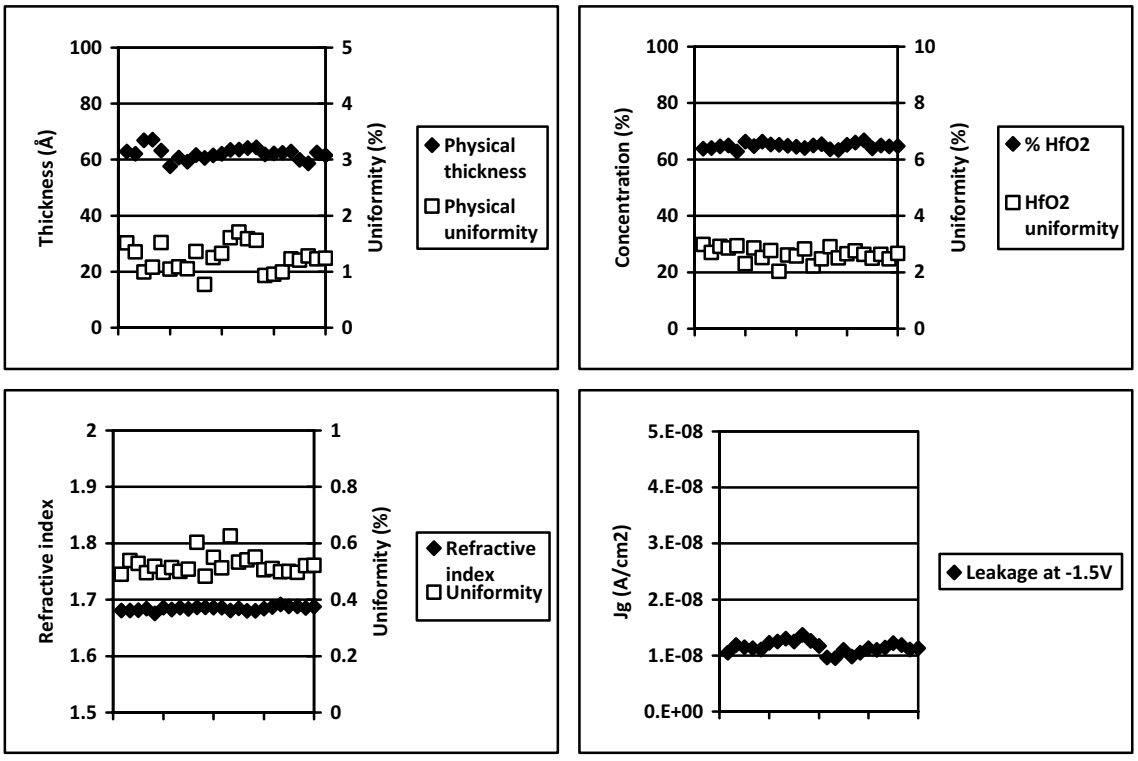


Figure 2. Results from all DOE and repeatability tests showing excellent process repeatability of co-injection ALD $Hf_{0.65}Si_{0.35}O_2$.

The results from the 6 repeatability wafers performed in the DOE were analyzed to evaluate the performance of the baseline process. The mean physical film thickness was 62.4 Å with all wafers under 1.61 % within-wafer uniformity and a wafer-to-wafer uniformity of 1.20 %. The average Hf composition was 64.4 % with all wafers under 2.98% within-wafer uniformity and a wafer-to-wafer uniformity of 0.90 %. The mean refractive index was 1.684 with all wafers under 0.54 % within-wafer uniformity and a wafer-to-wafer uniformity of 0.17 %; and the leakage current ranged from 9.6×10^{-9} to 1.2×10^{-8} A/cm^2 measured at -1.5 V.

By comparing the 2 datasets it is clear that even with a >10 % change in stage temperature and 20 % change in precursor pulse time and chamber pressure the co-injected ALD $Hf_{0.65}Si_{0.35}O_2$ demonstrated excellent process stability. These process parameter changes are extreme relative to the current hardware control.

Long term repeatability of films with a physical thickness of 20 Å deposited with the same baseline process has been confirmed over several months including periods of preventative maintenance and has yielded wafer-to-wafer thickness variation of <1.5 % with all individual wafers maintaining <2.0 % within-wafer uniformity.

Statistical analysis of these DOE process results was performed using JMP software. Based on the collected data the software produced a series of prediction profiles for each of the measured parameters showing their sensitivity to the substrate temperature, precursor pulse time and chamber pressure. Physical thickness increased with all three process parameters varied in the DOE as shown in Figure 3. The same trends were observed from the F5x optical thickness results, but with about a 14 Å offset corresponding to the interfacial oxide layer on the substrate.

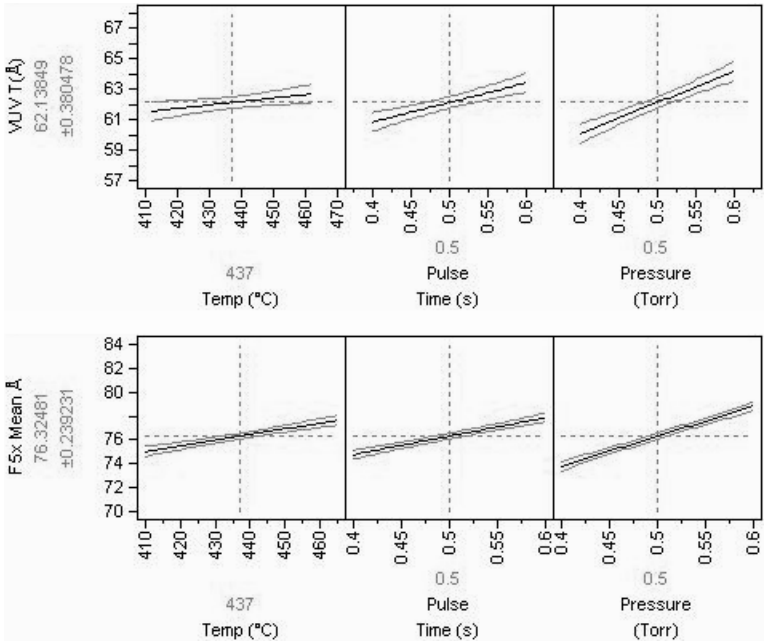


Figure 3. JMP predictions for $Hf_{0.65}Si_{0.35}O_2$ physical and optical film thickness.

The refractive index increased with temperature, indicating that the Hf concentration is increasing. The VUV Hf composition measurements verify this trend (Figure 4), associated with the lower reaction temperature of the TDMAHf precursor relative to the 3DMASi precursor. Pulse time had a minimal impact on the refractive index and Hf composition. This can be explained by the co-injection technique delivering both

precursors to the wafer surface simultaneously at controlled relative concentrations under conditions yielding a reasonably self-limiting saturated monolayer reaction. Higher chamber pressure decreased both the index and the Hf concentration. As the 3DMASi is less reactive than the TDMAHf, greater chamber pressure allows the 3DMASi to reside on the wafer surface longer and usurp some of the sites that TDMAHf would otherwise occupy.

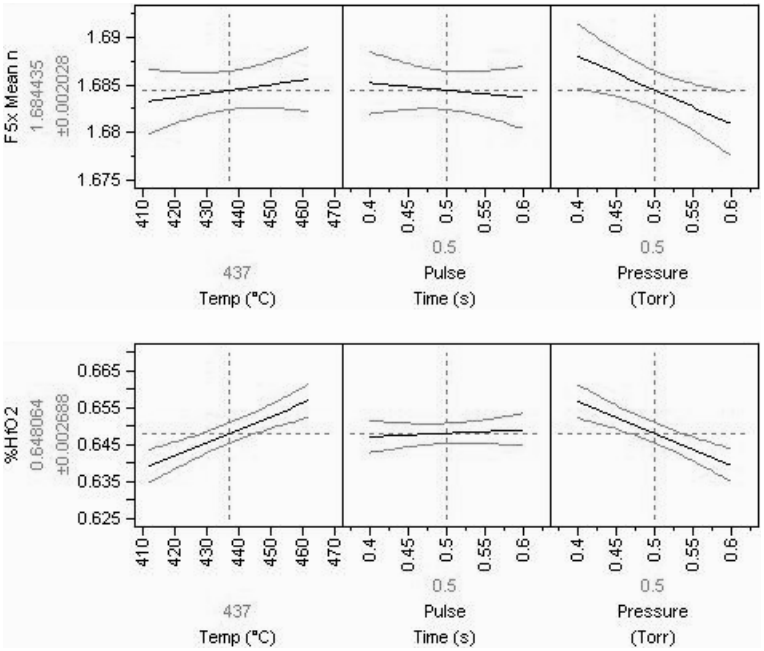


Figure 4. JMP predictions for $Hf_{0.65}Si_{0.35}O_2$ RI and Hf concentration.

The leakage current shows a decreasing trend with higher pressure as shown in Figure 5. It is believed that this can be explained from the increasing film thickness and reduced hafnium content at higher process pressure.

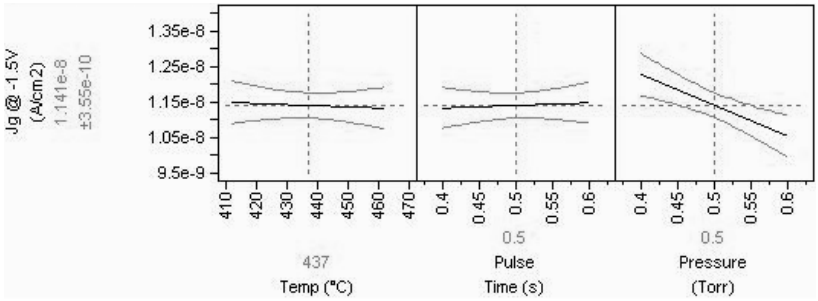


Figure 5. JMP predictions for $Hf_{0.65}Si_{0.35}O_2$ leakage current.

In a manufacturing environment, the robust nature of the co-injected ALD $Hf_{0.65}Si_{0.35}O_2$ has significant benefits in maintaining a repeatable high-yield process. However, it is important that the process also offer the flexibility required to meet the specific physical and electrical properties for successful device integration. It has been demonstrated that the co-injection ALD process is flexible enough to deposit $Hf_xSi_{1-x}O_2$ films with a Hf concentration from 30 to >80% (5). Figure 6 shows the leakage current density and Capacitive Equivalent Thickness (CET) obtained from a device employing a $Hf_xSi_{1-x}O_2$ gate dielectric deposited by the same technique. The data shows that the co-

injection process also enables excellent electrical properties to be obtained following device integration with a wide range of Hf concentrations.

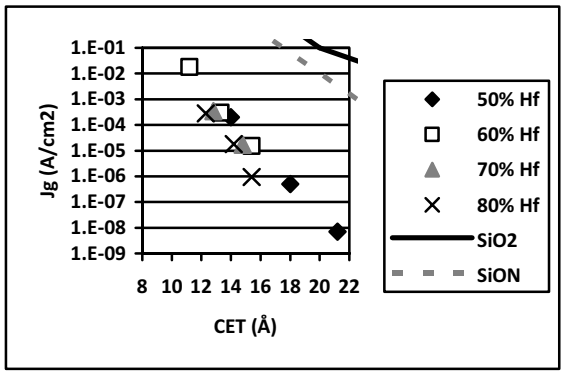


Figure 6. Electrical data obtained from a device employing co-injected ALD $Hf_xSi_{1-x}O_2$ gate dielectric demonstrating low Jg and CET over a wide range of Hf compositions.

Conclusions

ALD $Hf_{0.65}Si_{0.35}O_2$ films deposited using co-injected TDMAHf and 3DMASi with O_3 have excellent process repeatability and a wide process window. Extreme variations of stage temperature by >10 % and pulse time and pressure by 20 % had only a minor influence on the physical and electrical properties of the $Hf_{0.65}Si_{0.35}O_2$ films. The 3 input parameters had the greatest influence on the film thickness, with a wafer-to-wafer uniformity of 3.57 %, as all had a positive trend.

Baseline depositions of 60 Å films have a within-wafer thickness uniformity of <2.0 % and have a wafer-to-wafer uniformity of <1.5 %. At 65 % Hf compositional uniformity <3.0 % within-wafer and <1.0 % wafer-to-wafer was achieved. Leakage current varied from 9.6×10^{-9} to 1.2×10^{-8} A/cm^2 measured at -1.5 V.

JMP analysis of the DOE results showed that stage temperature has a weak, positive effect on $Hf_{0.65}Si_{0.35}O_2$ film thickness, Hf concentration and RI. The increase in Hf concentration is believed to be associated with the lower reaction temperature of the TDMAHf relative to the 3DMASi. Precursor pulse time had a weak, positive effect on film thickness only, largely due to the co-injection technique delivering both precursors to the wafer surface simultaneously. Pressure had a positive effect on thickness and a negative impact on Hf concentration, RI and leakage current. The effect on leakage current is likely caused by the increasing film thickness and reduced hafnium content due to the increased reactivity of 3DMASi at high pressure.

ALD $Hf_xSi_{1-x}O_2$ gate dielectrics deposited by the same co-injection technique demonstrated excellent electrical properties following device integration with a wide range of Hf concentrations.

Acknowledgements

The authors would like to thank our valued customers for providing the electrical data of the integrated $Hf_xSi_{1-x}O_2$ gate dielectrics.

Ex. 1016, Page 650

References

1. S. Thomson, P. Packan, M. Bohr, *MOS Scaling - Transistor Challenges for the 21st Century*, Intel Technology Journal Q3' 1998
2. K. Chang, J. Shallenberger , F-M Chang, K. Shanmugasundaram, R. Roman, P. Mumbauer and J. Ruzyllo, *Characteristics of Hf(Si,O) gate dielectrics as a function of Hf content*, International symposium on advanced gate stack, source/drain and channel engineering for Si-based CMOS, 2005 , vol. 5, p. 404
3. M. Houssa, M. Aoulaiche, S. Van Elshocht, S. De Gendt, G. Groeseneken, and M. M. Heyns, *Impact of Hf content on negative bias temperature instabilities in HfSiON-based gate stacks*, Appl. Phys. Lett. 86, 173509 (2005)
4. M. J. H. van Dal, G. Vellianitis, R. Duffy, B. J. Pawlak, L-S Lai, A. Hikavyy, N. Collaert, M. Jurczak and R. J. P. Lander, *Material aspects and challenges for SOI FinFET integration*, 213th ECS Meeting, 2008
5. L. Bartholomew, T. Lazerand, V. Rao, A. Bavin, M. Weldon, L. Wang, *Advanced Process Control of $Hf_xSi_{1-x}O_2$ Composition using Co-injection Atomic Layer Deposition*, ISTC/CSTIC 2009

ECS Transactions, 18 (1) 625-634 (2009)
10.1149/1.3096512 © The Electrochemical Society

Plasma Charging Damage Affecting Flash Memory Device Encountered at Pad Etch

S. F. Chong, *K. A. Mohammad*, F.S. Ng, L.J. Liu, D.G. Lee, and S. F. Lee

X-FAB Sarawak Sdn. Bhd., 1 Silicon Drive, Sama Jaya Free Industrial Zone, 93350 Kuching, Sarawak, MALAYSIA

This paper presents a study of plasma charging damage impact to 0.18 micron flash memory product. The affected product encountered high voltage N-MOS transistor leakage current. Correlation analysis pointed towards the passivation/ pad etch step. Engineering splits were performed to identify the potential root cause including non-optimized recipe and defective consumable kit. Further splits were performed with the consumable kits divided into three groups; the quartz group, the anodized aluminium group and the gas distribution plate group. The focus of this paper is to highlight the combined impact of high radio frequency (RF) power from process recipe and intrinsic defect from consumable parts. In addition, this paper also investigates the effectiveness of electrical test rework procedure utilizing an ultraviolet (UV) erase.

Introduction

Plasma charging damage has been reported by Toshiba as early as 1983 with batch processing gate oxide [1]. McVittie *et. al.* reported that the charging could be related to plasma conditions, oxide quality and device structure [2]. Cheung *et. al.* defined plasma charging damage as the unintended high field stressing of gate-oxide in MOSFET during plasma processing [3]. Cheung *et. al.* also highlighted that the stress voltage developed across the gate and substrate during plasma processing comes from three major sources, i.e., a non-uniformed distribution of plasma potential across wafer, the charge filtering due to microscopic topography of the wafer (shading), and AC effects due to the nature of radio frequency (RF) discharge sustaining the plasma [4].

Charging Effect

Plasma charging effect has introduced many serious process damages in term of profile [5] as reported by Hwang *et. al.* Most researchers agreed that the charging damage is caused by the Fowler-Nordheim (F-N) current injected into the gate oxide during plasma processing [6-9]. Extensive research has been carried out to minimize the charging effect. Steiner *et. al.* proposed a novel method to perform wafer de-chucking using power-lift to minimized this type of damage [10]. Barlingay *et. al.* investigated the interaction of ultraviolet (UV) radiation with plasma charging and proposed a method to recover the damaged device using UV erase and bake cycles [11]. Yue *et. al.* performed analytical studies including the high RF power in dielectric etch, the process kit, and the wafer grounding. He conclued that a tighter process kit specification control with etch rate improvement could minimize the current charging damage significantly [12].

Plasma Charging Damage at Pad Etch

In this paper, plasma charging damage was reported at E-Test, as shown in Fig. 1. The root cause finding later pointed to Pad Etch.

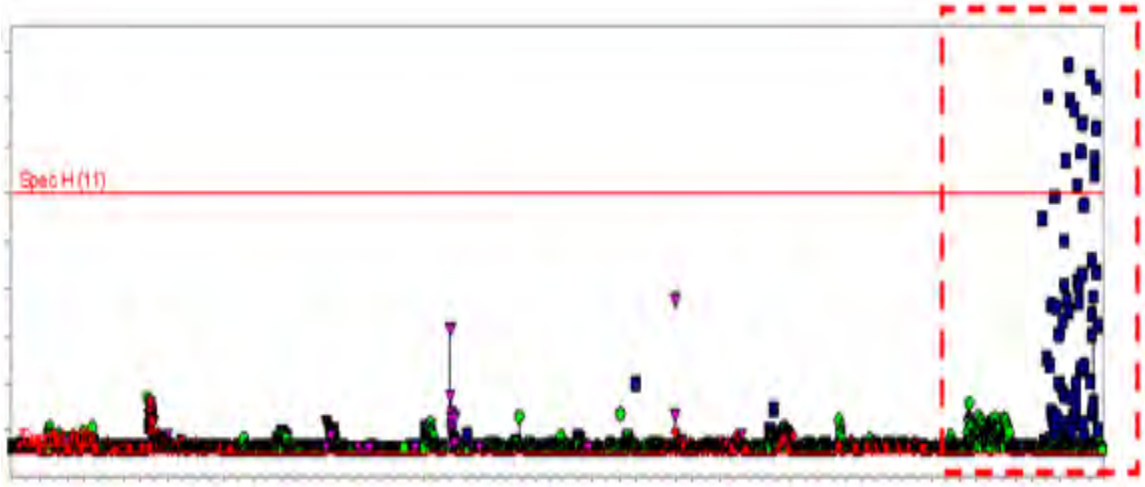
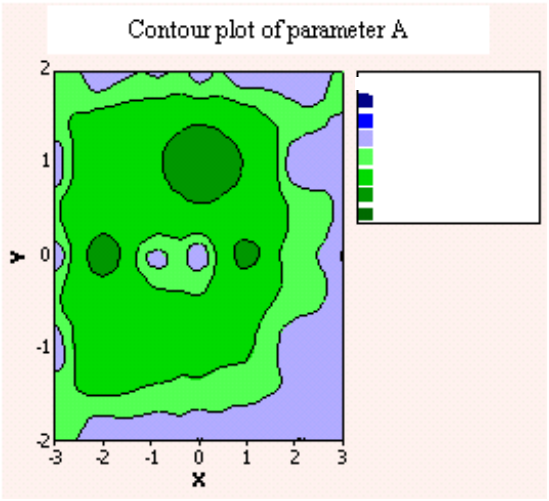
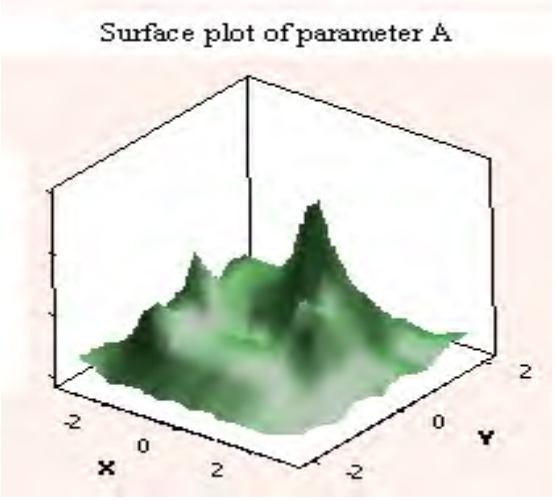


Fig. 1 The trend chart showing the affected lot failing the electrical test (e-test) parameter A

This electrical test parameter is related to current leakage at N-channel type Metal-oxide-semiconductor (NMOS). The full wafer map of the corresponding bad wafer, including a contour plot and a surface plot in Fig. 2 below, reveals the failed sites locating mainly at the centre region of the wafer.



(a) Contour Plot of Bad Wafer



(b) Surface Plot of Bad Wafer

Fig. 2. Full map e-test result for parameter A

Experiments were constructed to investigate these failed sites at the centre of the wafer map. The corresponding experimental parameters include the etch rate trend for one cycle, the process of record (POR) recipe, the affected chamber, and the process kit used during that problematic period.

Ex. 1016, Page 654

Experiment & Simulation

The etch rate trend study using the POR recipe for one preventive maintenance (PM) cycle on the affected pad-etch chamber was performed. Results, presented in Fig. 3 below, showed evidently that the etch rate of the centre area is higher than that of the edge throughout the PM cycle. This also matched the trend of the e-test full map shown in Fig. 2. However, this was not confirmed as the root cause.

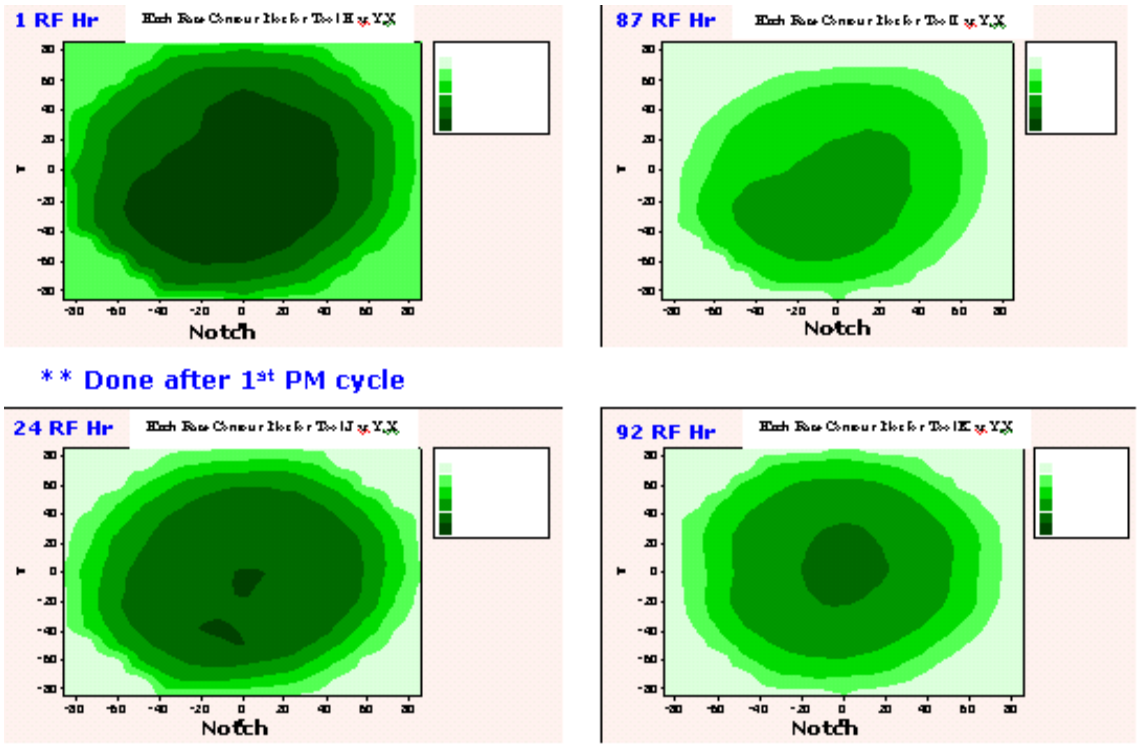


Fig. 3 Etch rate for the affected chamber during Preventive Maintenance (PM) Cycle using the POR recipe.

Plasma Damage Source

Recipe comparison and process kit segregation was performed to isolate the problematic process kit. Full stack engineering wafers were launched to investigate the actual root cause. The engineering wafers, consisting of 3 Metal layers, were prepared by skipping defect and measurement steps except for critical key area.

POR Recipe

The POR recipe employed has a highly-polymerized nature due to the use of a high radio frequency (RF) power of *b* watt (W) and large gas ratio (CHF_3/CF_4). The passivation stack to be etched away was >17,000 Å, which consists of high density plasma (HDP) oxide and passivation silicone nitride (SiN) as shown in Fig. 4. The time to etch the passivation stack, *t* including top Titanium Nitride (TiN) is

$$t = \text{A seconds (HDP)} + \text{B seconds (SiN)} + \text{C seconds (TiN)}, \text{ where } C > B > A \quad (1)$$

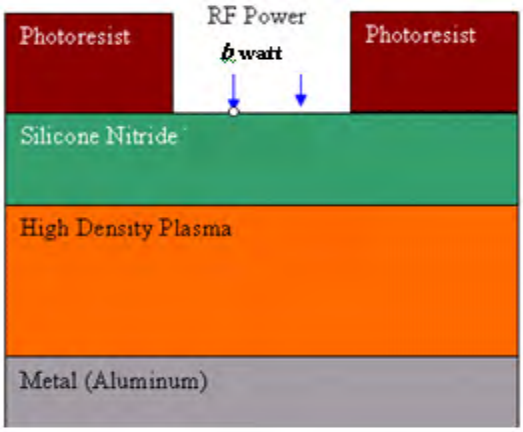
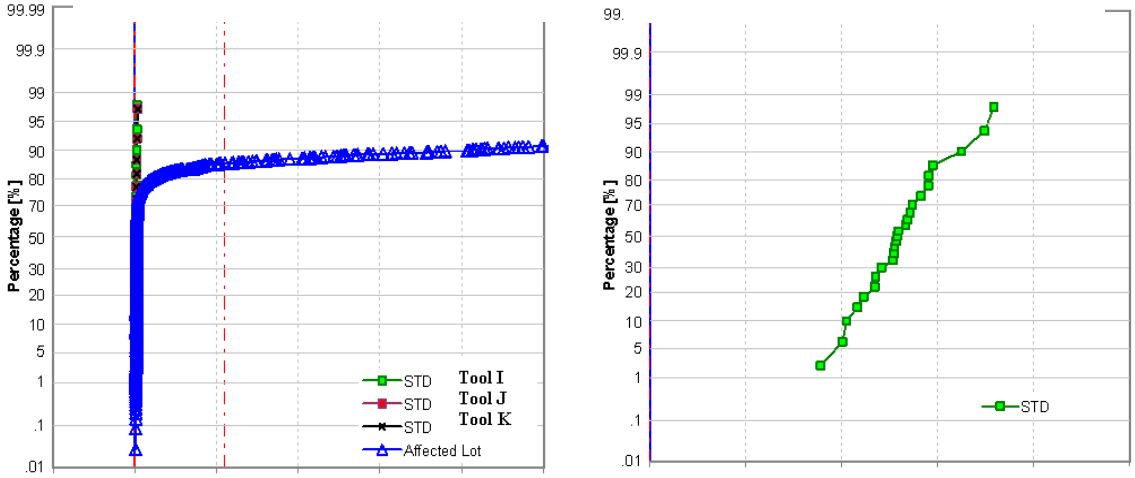


Fig. 4 The passivation stack showing the high RF Power being employed to remove passivation layer until Aluminum is exposed.

The high RF power recipe combined with longer processing time is one of the major factors causing electrical parameter to shift, as shown in Fig. 5. Longer processing time tends to cause more metal to be exposed to physical bombardment. Apart from this, we also suspect that the metal pad could act as an antenna *absorbing* the charges from the plasma since the charging seems also to be electrical test structure dependent. Depending on the gate oxide thickness, a high magnetic/Gaussian power and longer processing time recipe can cause gate oxide breakdown if not optimized. In addition, high RF power is also renowned for causing arcing damage.



(a) Shifted parameter A compared to standard

(b) Zoom in for parameter A for standard

Figure 5. Cumulative plot for E-Test parameter A.

Part of the focus of this paper is to study the recipe effect on parameter A. The affected lots have parameter A values greater then zero pA/µm from E-Test, as shown in Figure 5.

Process kit segregation

Based on the observation that the problem ceased to exist after a wet clean and PM, the process kit used during that cycle was isolated and investigated.

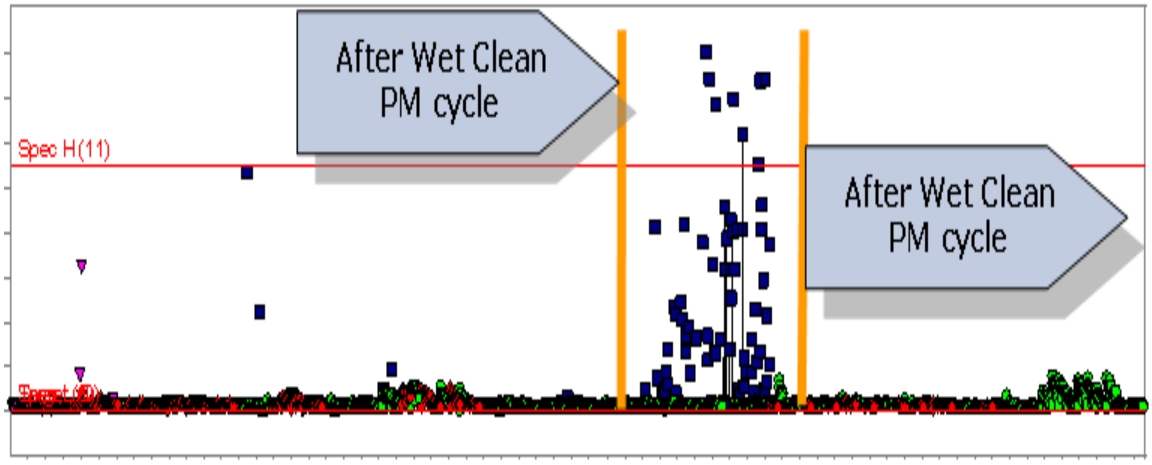


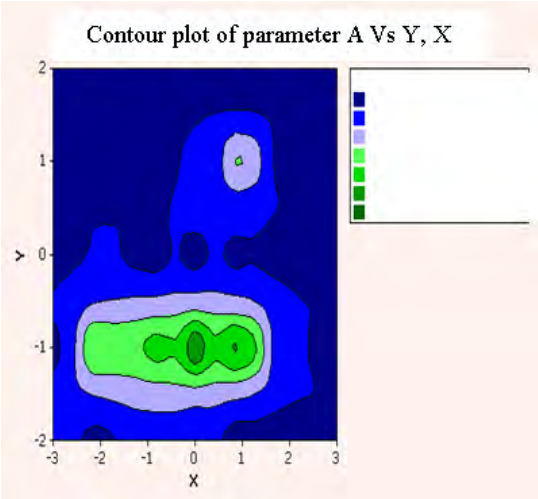
Fig. 6 The trend for E-Test parameter A grouped by process chambers.

The process kit was grouped into three, as shown in Fig. 7. The anodized coating group consists of mainly anodized aluminum kit; i.e. chamber liner, cathode liner, slit liner door and separation ring. The gas distribution plate is the major kit for the gas delivery inside the process chamber. The third group consists of quartz and silicon parts which were replaced during wet clean PM. The 'Pass' result is determined by passing all e-test points or full map for these three wafers for each group.

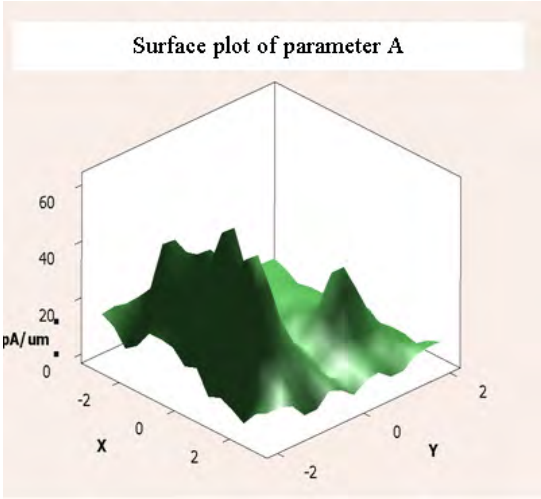
Group	Part	Result
1	Anodize Coating Parts: - Chamber liner - Cathode liner - Slit Liner door - Separation Ring	Pass
2	- Gas Distribution Plate (GDP)	Pass
3	Quartz and Silicon Parts: - Quartz Insulator - Shadow Ring - Silicon Insert Ring - EPD Window	Fail

Fig. 7 Process kit root cause finding split by group.

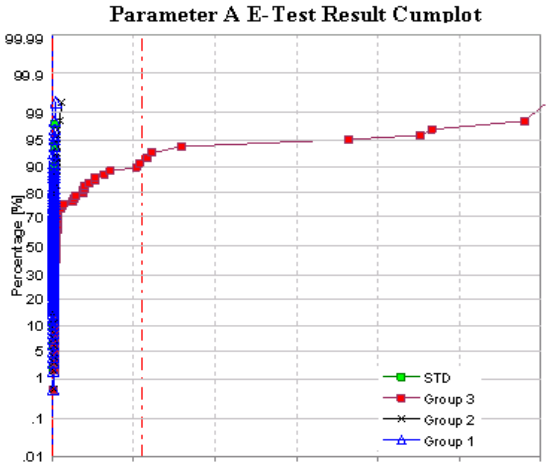
The process kit was monitored using 3 wafers for each run to verify the failed or bad process kits group. Results showed that Group 3 failed the E-Test, as shown in Fig. 8.



(a) Contour plot for Group 3 E-Test Parameter A Result



(b) Surface plot for Group 3 E-Test Parameter A Result



(c) Cumulative plot for Process Kits E-Test Parameter A result

Fig. 8. Process Kit Simulation Result

This trend is similar to that observed in Fig. 2, which confirmed that the kit from Group 3 is problematic and is part of the root causes contributing the charging issue.

Process flow change

There was a change in the product process flow, which involved relocating the Pad Alloy step before the pad passivation step. The histogram is shown in Fig. 9. There were intermittent spikes for the data from upper spec value of *l* pA/µm before the anneal step change. The change was suspected to cause excess hydrogen (H_2) during annealing which is related to memory data retention. This gave us the initiative to perform rework for the failed wafers using an ultraviolet (UV) erase before verified by the corresponding electrical test. The experiment was carried out with 1 wafer as pilot. The success criteria were the same as those of others, i.e., passing all standard E-Test and Full map for parameter A.

Ex. 1016, Page 658

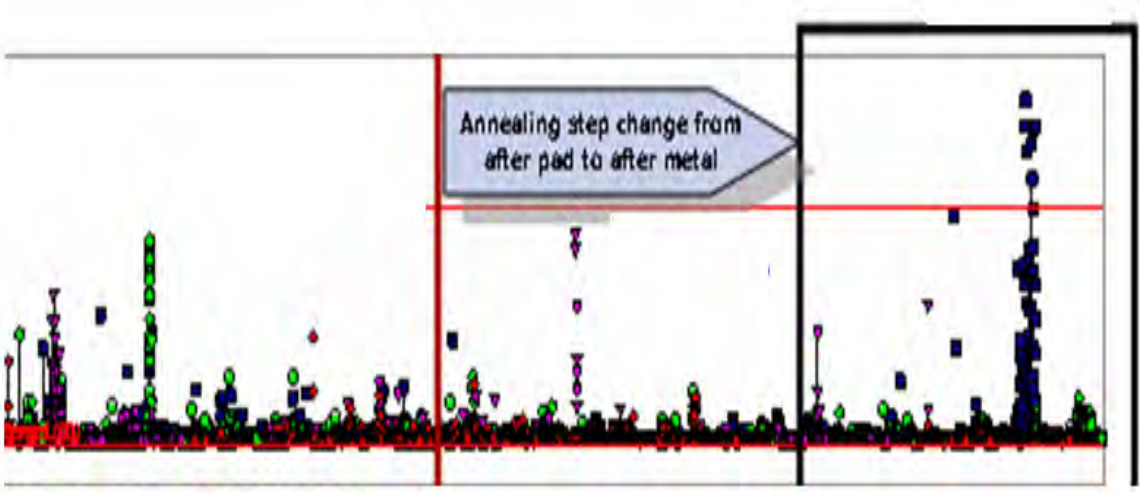


Fig. 9 The histogram of E-Test Parameter A due to the process flow change

The electrical rework results, shown in Fig. 10, revealed that wafers using UV erase is damage free without physical impacts on the pad structure or on metal grain structure, which is important for device performance. The wafer was sent for failure analysis to compare the metal grain structure and the structure of the passivation stack after rework with the standard wafer. Results of the failure analysis using scanning electron microscope (SEM) are shown in Fig. 11. The centre and edge of the bond pad area was compared to find no abnormality. The metal grain structure was almost similar to the standard wafer while the passivation structure was still in tact.

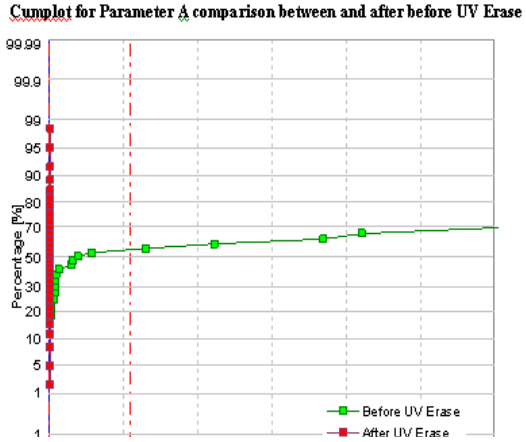


Fig. 10 Electrical rework result for E-Test Parameter A

The UV erase, being able to remove excess trapped electron that causes current leakage for parameter A momentarily, is only a temporally containment. The long term solution is to develop a new low power recipe for pad etch.

Ex. 1016, Page 659

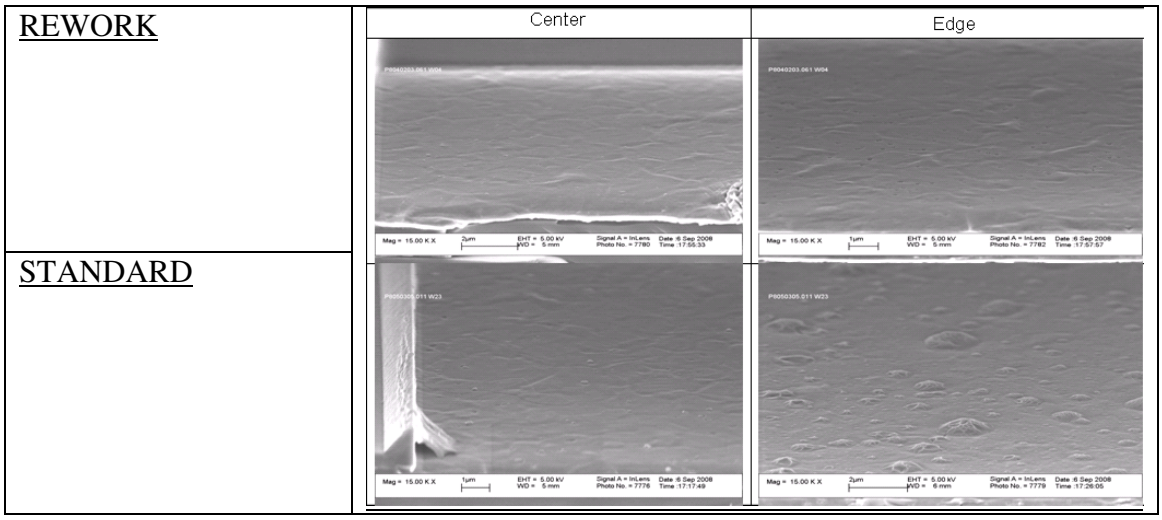


Fig. 11 SEM Image of Failure Analysis for Rework and Standard Wafer at Pad Open

The resulting new recipe has the key steps below, as shown in Fig. 12.

- The first step was to remove 80% of top passivation nitride (SiN) and oxide (HDP) layer with high RF Power *p* W (1100 W – 1800 W) [Physical Etch Dominant]
- The second step was to employ soft landing with lower RF Power *q* W (800W – 1000 W) to prevent plasma charging damage on exposed metal.
- The third step was to employ chemical etch to break TiN and exposed metal line with Low RF Power *r* W (400 W – 550 W). [Chemical Etch Dominant]

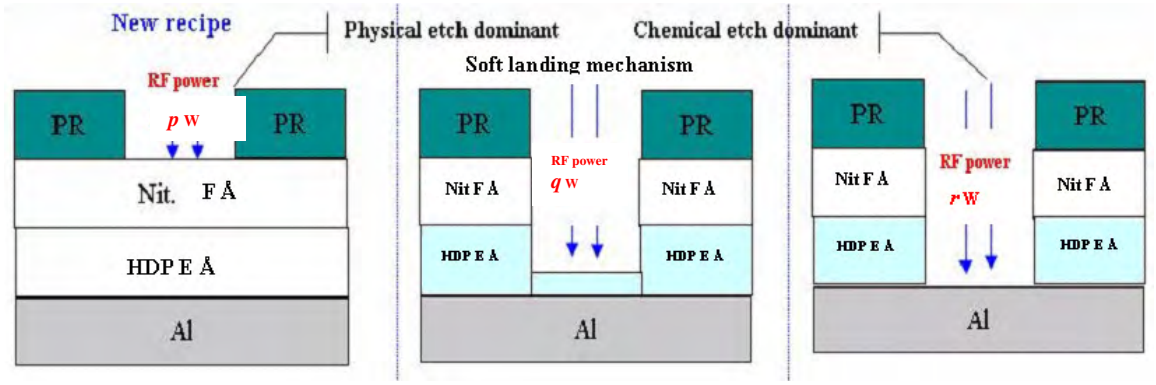
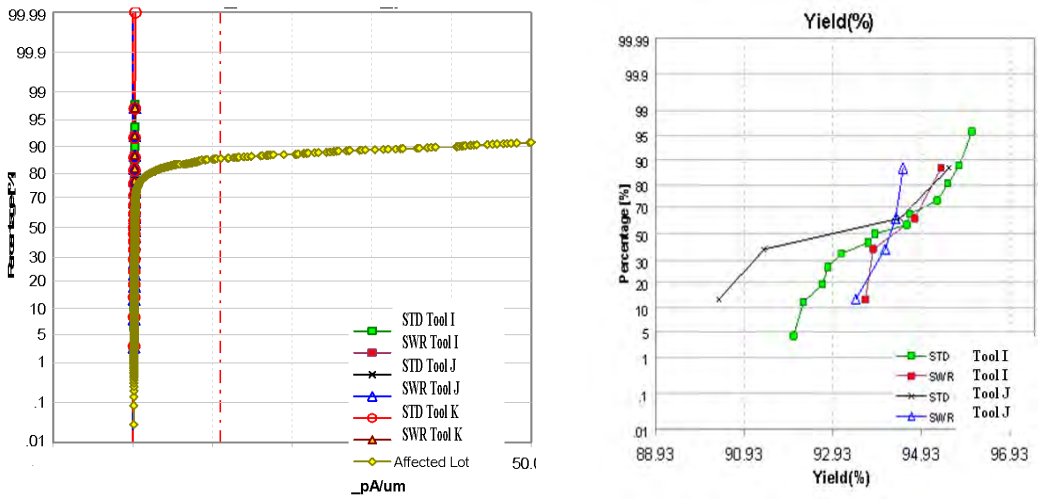


Fig. 12 Key steps for the new low power pad etch recipe

Figure 13(a) shows that the SWR and STD pad recipe are comparable in term of e-test performance. The corresponding Sort results are also promising, where the SWR yielded a 2% improvement, as shown in Fig. 13 (b). Based on the Sort 1 result, the SWR Low Power recipe is optimized in a way to have minimized the plasma charging damage.



(a) Cumulative plot E-Test result for SWR and STD wafers for parameter A (b) Sort yield result for SWR and STD wafers

Fig. 13 Comparison between Standard (STD) and Low Power Pad Etch (SWR) Recipe.

The optical inspection, manually performed to compare between STD and SWR wafers, showed no abnormality on the pad open area, as shown in Fig. 14 below.

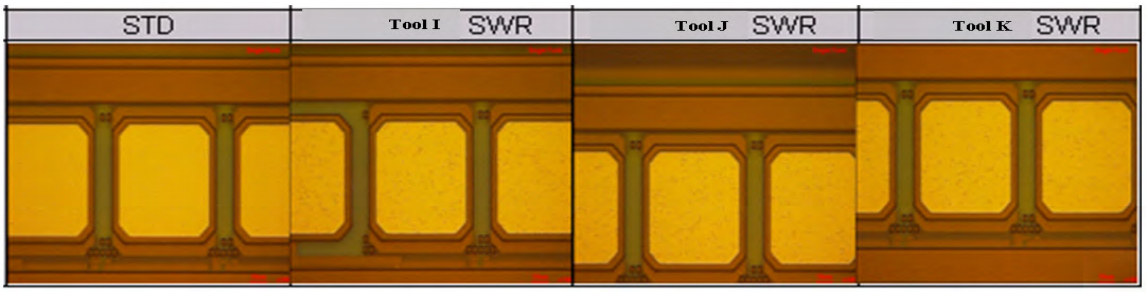


Fig. 14 Optical inspection data between standard wafer and low power recipe

Conclusion

We investigated the plasma charging damage encountered at pad etch. The affected device (flash memory) experienced excessive current leakage, as shown by e-test parameter A. We conclude that the root causes are the followings:

a) The defective quartz and silicon process kits during plasma processing.
b) The current POR process recipe employing high RF power for etching, which utilizes physical bombardment until metal is exposed causing plasma charging damage.

The new recipe, using soft landing mechanism with reduced RF power and chemical etching during metal expose to minimize charging damage, has been shown to be very effective to remove the plasma charging issue. Minimum parameter A issues were identified after this new pad etch recipe was released to production. The electrical rework using UV erase was implemented in a way that any product found with e-test failed parameter would be automatically triggered for e-test rework and then sent for re-testing.

Ex. 1016, Page 661

Acknowledgments

We would express our gratitude to the X-FAB Management Team for sponsoring and supporting this paper. We would also like to extend our humble appreciation to Mr. Andrew Paul, Mr. Thomas W. Suip and Miss Florence Chai from Yield Enhancement and Defect Management for their advice and technical contributions to this paper.

References

1. Y.Yoshida and T.Watanabe, Proce. Symp. Dry Process. pp.4, Tokyo (1983).
2. J. P. McVittie, 1st Symp. Plasma Process Induced Damage (1996).
3. Kin. P. Cheung, *Plasma Charging Damage,* Springer-Verlag, London (2000).
4. S. Fang and J. McVittie, *IEEE Electron Devices Letter,* Vol. 13 No. 5, pp. 288-290, (1992).
5. G. Hwang and K.Giapis*, J. Vacuum Science and Technology B*, **15**, pp.70–87, (1997).
6. K. Tsunokuni, K. Nojiri, S. Kuboshima and K. Hirobe: Ext Abstr SSDM p.195, (1987).
7. H. Shin, C-C., King and C. Hu: Proc IRPS, p.37, (1992).
8. K. Hashimoto, D. Matsunaga and M. Kanazawa: Proc. 13th Symp. Dry Process, p. 93, (1991).
9. S. Fang and J. P. McVittie: J. Appl. Phys. **72,** (1992).
10. K. G. Steiner and K.P. Cheung, 6th Symp. of Plasma Process Induc Damage, (2001).
11. C. K. Barlingay, R. Yach and W. Lukaszek, 7th Symp. of Plasma Process Induc Damage, (2002).
12. K. H. Yue, C. T. Tay, K. L. Se and B. K. Goh, Proc. 2008 1EEE/SEMI ASMC, pp. 282-286, (2008).

ECS Transactions, 18 (1) 635-639 (2009)
10.1149/1.3096513 © The Electrochemical Society

Material Etch Rate Control in the Fluoride Containing Stripper for Post Etch and Ashing Residue Removal

Bing Liu, Libbert Peng, Julia Peng, Joey Yu, and Shumin Wang

Anji Microelectronics (shanghai) Co., Ltd., Shanghai 201203, China

In this paper, we share a unique metal and dielectric inhibitor system and the corresponding methodologies in controlling the materials etch rate from the fluoride-containing stripper and the etch rate stability with the bath life time extension. In addition, the cleaning performance of the fluoride containing stripper with low materials etch rate is also introduced.

Introduction

Fluoride containing stripper has been used as a post etch and ashing residue removal solution since mid-1990s due to its cost-effective and unique technical advantages, such as intermediate solvent rinse elimination and oxide surface skimming-off [1]. However, there are still many challenges to be addressed, such as the undesirable high dielectric materials etch rate, weakness to organic residue removal, and limited performance in via and bond pad cleaning. Therefore controlling the metal and dielectric etch rate stability throughout the bath life, as well as providing the robust cleaning, is a key success indicator for a stripper solution. Although leading edge logic chip manufacturers has made the conversion from AlCu line to copper damascene and low K device at 130nm technology and beyond, AlCu for DRAM, CIS and hence the aluminum-based process technology still plays an important role in IC manufacturing industry, especially in China [2-4]. In this work, we have demonstrated how to control materials etch rate in the fluoride stripper system with a focus on simultaneously the post etch and ashing cleaning in the AlCu technology. Finally, the cleaning performance of the fluoride stripper is also introduced.

Experimental

Extensive metal, via, pad and photoresist wafers are used in the evaluations. Metal film stack consists of Ti/TiN/AlCu/Ti/TiN/Oxide, via film stack TEOS/HDP/Ti/TiN/AlCu and Pad film stack SiN/TEOS/HDP/Ti/TiN/AlCu. All of them are post etch/ash wafers without any chemical treatment. Photoresist wafers contain a 2um g-line positive photoresist with soft baking at 90℃ for 120s. Process equipment used in this study is a Semitool SST for full wafer evaluation and for temperature-control bath of coupon wafers. The corresponding analytical equipments include a 4-point probe RS-75 for metal etch rate measurement, a Nanospec 6100 for dielectric etch rate measurement, and a Hitachi 4700 for the evaluation of cross section cleaning performance.

Ex. 1016, Page 663

Results and Discussions

Metal Etch Rate Control

A fluoride stripper system generally contains the following main components, i.e., the solvent, the de-ionized (DI) water, and the fluoride source, and may include certain additional components such as chelator, inhibitor, surfactant and other special chemical agents. One key factor for materials etch rate control is the optimization of inhibitor system. Various common metals, including AlCu, Cu, Ti and W, are encountered in the AlCu technology BEOL (back end of line) process. We used AlCu to serve as the vehicle for the performance evaluation of the newly developed metal inhibitor. The corresponding details are listed in Table 1.

TABLE 1 AlCu etch rate control by the unique metal inhibitor in the fluoride system

Formulation No.	Fluoride (%)	Inhibitor (%)	AlCu etch rate at 25℃, A/min
1	w/o	0	1.46
2	w	0	3.09
3	w	Inhibitor A, 0.1	1.88
4	w	Inhibitor B, 0.1	1.02

Other components are the same.

Table 1 shows that Formulation 1 has relatively low AlCu etch rate without the fluoride source and the AlCu etch rate increases with the fluoride addition. Both inhibitor A and B show good inhibition to AlCu corrosion and the latter is the better of the two. Based on this information, two kinds of solutions were further developed. The corresponding detailed metal etch rates are listed in Table 2.

TABLE 2 Different metal etch rate obtained from optimizing the fluoride system

Formulation	different metal etch rate, A/min			
	Al	Cu	Ti	W
IDEAL Clean™ solution A, at 35℃	0.92	1.54	0.32	1.12
IDEAL Clean™ solution B, at 45℃	0.18	2.16	0.08	0.29

Table 2 shows that etch rates of AlCu, Cu, Ti and W are now all very low in the fluoride containing strippers, resulting in an adequate margin to prevent metal corrosion in post etch and ashing residue removal process.

Dielectric Etch Rate Control

In the fluoride containing stripper system, dielectric material loss mainly is caused by the attack from fluoride species. The fluoride species identified in the NH_4F system are shown in Figure 1. The etch rate for different dielectric materials is very species dependent. For example, SiO_2 etching is dominated by HF_2^- while BPSG etching is controlled by nondissociated (neutral) HF [5]. Therefore the key of minimizing materials etch rate is to optimize these various relevant fluorine-containing species. On the other hand, certain inhibitor addition may also be very helpful in controlling material loss.

Ex. 1016, Page 664

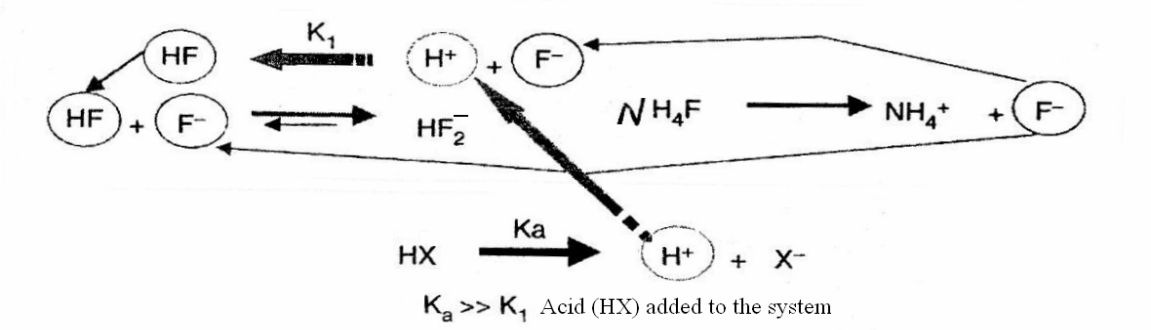


Fig. 1 Various fluoride species in the NH_4F system

In the AlCu technology BEOL process, the dielectric materials generally include HDP oxide, PE-TEOS, FSG, BPSG, and SiN. We have decided to use BPSG as vehicle to evaluate our unique inhibitor and the results are listed in Table 3.

TABLE 3. BPSG etch rate control by the unique inhibitor in the fluoride system

Formulation No.	Inhibitor (%)	BPSG etch rate at 40℃, A/min
1	0	8.41
2	Inhibitor C, 0.1	0.85

Other components are the same.

Table 3 reveals that Inhibitor C meets the BPSG etch rate control requirement. Based on this information, a solution is developed to be tested for other dielectrics. The resulted dielectric materials etch rates are listed in Table 4, which shows that the IDEAL Clean™ solution C newly developed can control the dielectric etch rate well at 40℃.

TABLE 4 Different dielectric etch rate in optimized fluoride system

Formulation	different dielectric etch rate, A/min		
	HDP oxide	PE-TEOS	BPSG
IDEAL Clean™ solution C, at 40℃	0.12	0.17	0.33

The Effect of Bath Life Extension on Etch Rate

In general, it is difficult to control the materials etch rate even in the fresh fluoride containing strippers. Moreover, it is a technical challenge to maintain the materials etch rate stability with the bath life extension, especially for the dielectric materials. Base on new solution (IDEAL Clean™ solution C) developed, stable materials etch rate has been achieved with bath life extension, as shown in Fig. 2.

Ex. 1016, Page 665

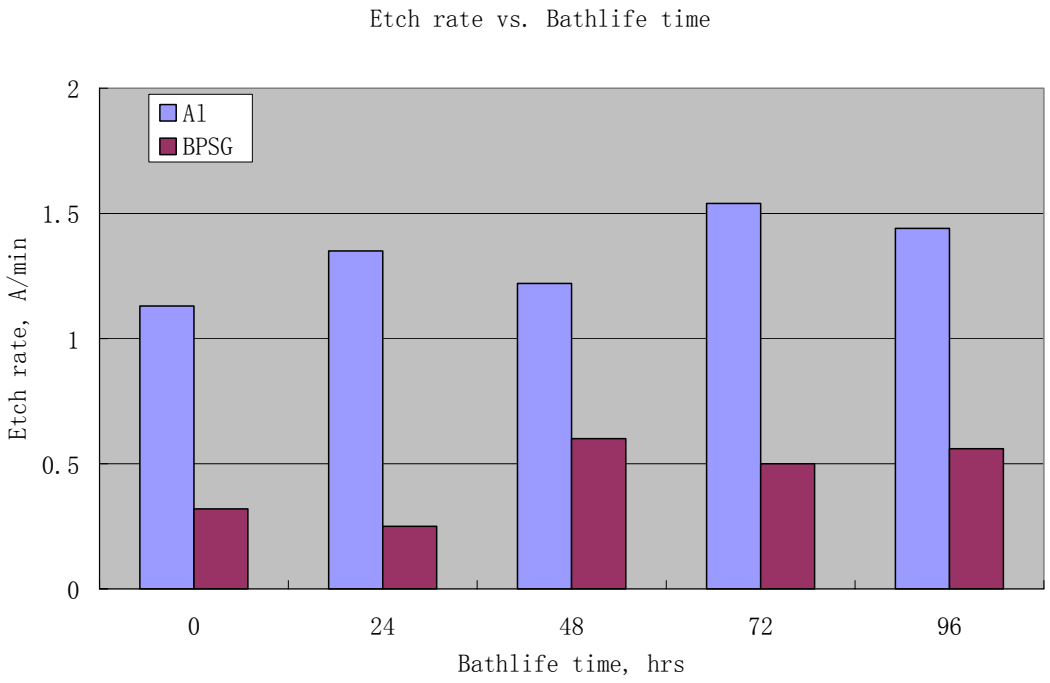
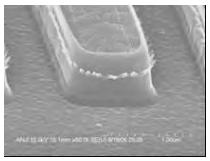
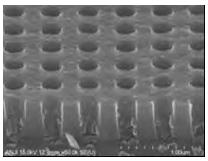
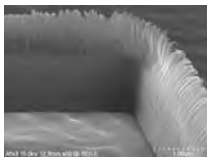
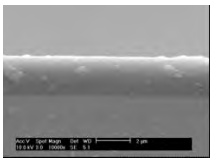
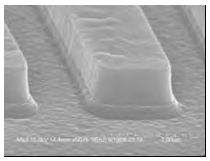
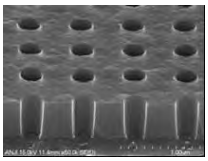
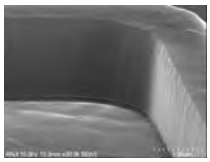


Fig. 2 Materials etch rates with bath life extension at 40℃ using IDEAL Clean™ solution C

Cleaning Performance

Controlling the materials etch rate is fundamental for post etch and ashing residue stripper, but it will not be useful if the polymer residue can not be completely removed in the process. The above mentioned IDEAL Clean™ solutions A, B and C can be used to clean metal, via, pad and photoresist wafer by using immersion or batch spray methods at different process time and process temperature. The process time varies from a few minutes to 30mins, and process temperature ranges from room temperature to 45℃ based on incoming wafers. Following the cleaning step is a DI water rinse and then a N_2 dry. The cleaning results of IDEAL Clean™ solutions C are summarized in the Table 5.

Table 5 The cleaning performance summary of IDEAL Clean™ solution C

	Metal	Via	Pad	photoresist
Controlled wafer				
Cleaning process	15 mins @40℃	30mins@40℃	20mins@40℃	15 mins @40℃
After cleaning				
Cleaning results	Clean/no corrosion & undercut	Clean/no enlarge & mushroom	clean	Clean

Summary

IDEAL Clean™series solutions are unique fluoride-containing strippers which has been demonstrated to have stable materials etch rate within bath life time and can be used for metal, via, and pad cleaning with solid performance. This overcomes the limitation of the existing conventional fluoride based system in the post etch and ashing residue removal process for the semiconductor AlCu technology.

Acknowledgments

The authors would like to thank ATMI for their kind support and Luck Zhang for high quality X-SEM picture collection.

References

1. Darryl Peters, Les Molnar and Rob Rovito, *Future Fab* International Issue 14 - WAFER PROCESSING
2. Libbert Peng**,** Bing Liu, Andrew Wang, David Yin, and Shumin Wang, *Semiconductor Technology*, P397 (2008).
3. Libbert Peng**,** Bing Liu, Yong Gong, and Shumin Wang, *9th International conference on Solid-State and Integrated-Circuit Technology* , E4.7 (2008).
4. Libbert Peng**,** Bing Liu, Shumin Wang, David Yin, and Todd Hogan, *Surface Preparation and Cleaning Conference*, (2008)
5. Tadahiro Ohmi, *Scientific Wet Process Technology for Innovative LSI/FPD Manufacturing*, P175 (2006)

Ex. 1016, Page 667

ECS Transactions, 18 (1) 641-643 (2009)
10.1149/1.3096514 © The Electrochemical Society

Influence of Polymeric Gas on Sidewall Profile and Defect Performance of Aluminum Metal Etch

Xin-Peng Wang[a], Hai-Yang Zhang[a], Man-Hua Shen[a], Wu Sun[a], and Shih-Mou Chang[a]

[a] Logic Technology Development Center, Semiconductor Manufacturing International Corporation, No.18 Zhang Jiang Rd., Pudong New Area, Shanghai, 201203, P.R.China

As the semiconductor industry moves to 65nm node and beyond, Al metal etch has become more challenging for its insufficient amount of photo-resist (PR) protection to achieve the desired sidewall profile and keep the corrosion free performance simultaneously. In this paper, we focus on the selecting appropriate polymeric gas and identifying the effect of various etch polymeric gases on Al metal sidewall profile and their corrosion defect performance. Three typical gases including N2, CHF3 and CH4 are selected in this paper. The observations from scanning electron micrographs indicate N2-based recipe is apt to form the severely tapering sidewall profile while CHF3-based recipe has loose polymer protection over the sidewall, thus leading to much worse corrosion defect performance. As expected, CH4-based recipe outperforms the two above polymeric gases. It not only shows the acceptable sidewall profile but also delivers the defect performance of free corrosion.

Introduction

Aluminum (Al) and Aluminum alloys have been extensively utilized as chip wiring materials in the fabrication processes of Cu interconnects. As the dimension of devices continues to scale down, Al metal tends to become thicker for the concern of adverse effect of high resistance connection on reliability. This poses the great challenge for Al metal etch process because it's tough for the insufficient amount of photo-resist (PR) to realize the desired sidewall profile and achieve free corrosion performance simultaneously (1~3). Such issue will exacerbate for higher transmission rate case. N2 and CHF3 are two typical polymeric gases widely used in Al metal etch. However, both of them could not fully satisfy the tightened process requirements caused by the inherent insufficiency of PR thickness. In this paper, we employed CH4 in Al metal dry etch process as the additive polymeric gas. The performance of all three polymeric gases is evaluated on the sidewall profile of Al metal line and the corresponding corrosion defect performance at the side-wall of Al-metal pad.

Experiment

Al metal etch was conducted on typical LAM-2300-Versys-Metal chamber. Three kinds of polymeric gas N2, CHF3 and CH4 were examined in this work. Their percentage was optimized based on our previous researches and ranges from 5%~10% in main etch and over-etch step, respectively. The same flow rate of Al metal main etch gas, Cl2 and BCl3, are applied for all three cases. In order to compare, we also employed the same source power and bias power in three polymeric gas splits. Detailed scaled Al metal etch

Ex. 1016, Page 669

recipes are shown in Table.1. The thickness of Al metal is 14um. The transmission rate of this Al metal etch process is about 85%. This means only 15% area is covered by photo resist before etch. Scanning electron microscope (SEM) was used to monitor the sidewall profile of Al-line. Defects performance was inspected after etch and wet clean process by KLA defect scan and review tool.

Table I. Summary of scaled Al metal etch recipes for different polymeric gas conditions

Split Condition	Polymerci Gas type	Etch Step	Pressure (mT)	Source power (W)	Bias power (W)	CL2 (sccm)	BCL3 (sccm)	Polymeric Gas usage (sccm)	Process time
1	N2	Main Etch	a	b	c	d	e	f	End point (151sec)
2	CHF3	Main Etch	a	b	c	d	e	f	End point (117sec)
3	CH4	Main Etch	a	b	c	d	e	f	End point (135sec)

Results and Discussion

Fig.1 shows the polymer performance at Al-pad side-wall and Al-line top view comparison under three different polymeric gases. Clearly, N2-based polymeric gas leads to more tapering profile. This can be attributed to that N2 gas brings heavier polymer during etch process and significantly slows down Al metal etching rate.This can be also noticed from that N2-based recipe features the longest endpoint in the main-etch of Al metal etch among three splits (Table.1). Longer etch time inevitably results in more photo

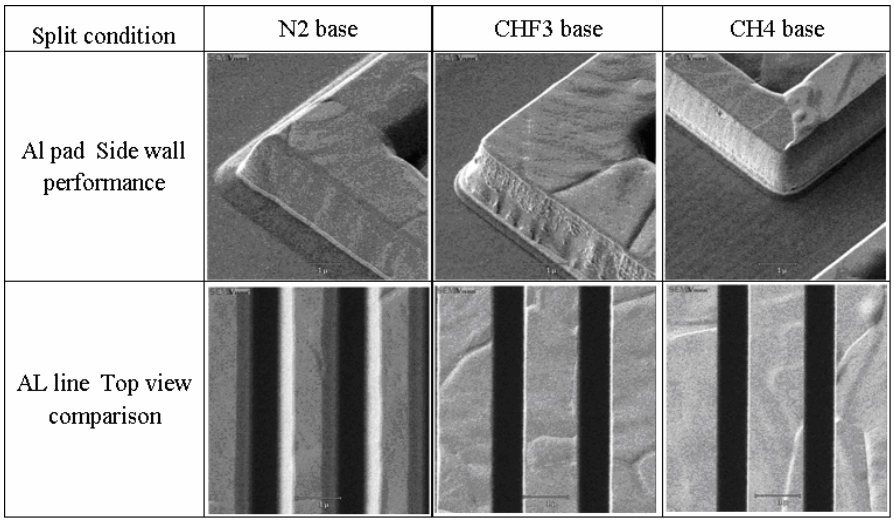


Fig.1 Al pad side wall and Al line top view comparison among three different polymeric gases

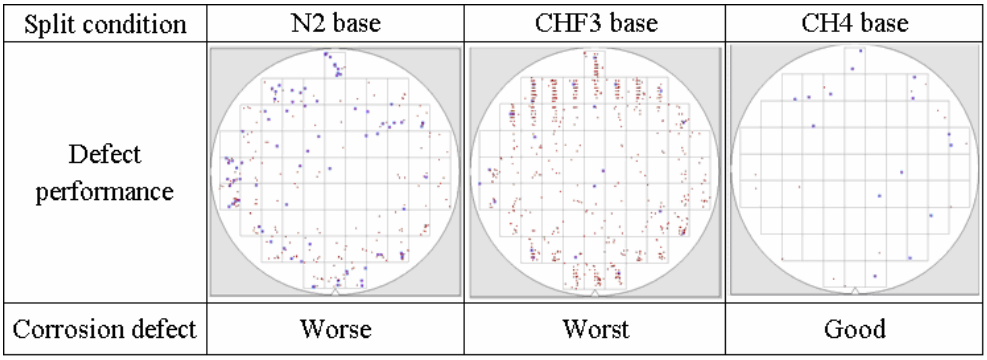


Fig.2 Corrosion defect comparison among three different polymeric gases

resist (PR) consumption. The remained PR cannot withstand physical bombardment in etch process. Coupled with heavy polymer, it finally caused much tapering profile of Al-pad and the top damage of Al line (trapezium top-profile). In contrast, CHF3-based split possesses the shortest end-point time of Al metal main etch. This indicates it's capable of generating much less polymer. In SEM picture, its roughest sidewall performance can be observed even with some Al metal grains exposed. As expected, CH4-based recipe provides the suitable polymer condition not only for Al-pad side-wall protection but also for the achievement of vertical Al-line profile.

The corresponding corrosion defect performance is summarized in Fig 2. As is well known (4-7), the corrosion defect performance can be ascribed to the side wall protection that is formed by polymer or by-product during Al metal etch process. Side wall protection mainly comes from the function of polymeric gas utilized in etch processes. Appropriate side wall protection can remarkably prevent Al metal from being attacked in the subsequent wet clean process and atmosphere environment, thus delivering the desired corrosion defect performance. In brief, CH4-based split is superior to other two kinds of polymeric gases from the point view of either Al-line profile or corrosion defect at Al-pad sidewall. While CHF3-based split demonstrates the worst corrosion defect because of its weak side wall protection.

Conclusions

In this paper, we investigated the impacts of the polymeric gas on the sidewall profile of Al-line and corrosion defect performance of Al-pad. Three typical polymeric gases include N2, CHF3 and CH4 are compared. The results show N2 gas causes the heaviest polymer, thus forming the severely tapering sidewall profile of Al-line while CHF3-based split provides the relatively loose polymer protection against the sidewall and leads to much worse corrosion defect performance. CH4-based split outperforms two other polymeric gases. It not only shows the vertical sidewall profile of Al-line but also delivers the defect performance of free corrosion at Al-pad.

Acknowledgments

The authors would to thank their colleagues at etch department of logic technology development center for technical assistance and useful discussions during the course of this work.

References

1. H. J. Lee, C. L. Hung, C. H. Leng, N. T. Lian, L. W. Young, T. Yang, K. C. Chen, and C. Y. Lu, *International Journal of Plasma Science and Engineering*, p.10, Volume 2008 (2008).
2. X. M. Bu, M. Luo, C. Y. Sin, F. Zhang, P. Yelehanka, T. J. Lee and A. Krishnamoorthy, *Microelectronic Engineering*, p.204, Volume 77 (2005).
3. S. C. Wu, Y. J. Lee, K.H. Chang, S. Y. Lin, A. T. Cho, C. F. Chiu, C. D. Huang and J. R. Sha, *Journal of Physics: Conference Series*, p.1256, Volume 61 (2007).
4. S. Albericia, D. Coulonb, P. Joubinb, Y. Mignotb, L. Oggionia, P. Petruzzaa, D. Piumia and L. Zanotti, *Microelectronic Engineering*, p.558, Volume 70 (2003).
5. G. Stojakovic and X. J. Ning, *Journal of Vacuum Science & Technology A*, p.1425, vol. 18 (2000).
6. J. Yamamoto, T. Hsieh and T. Kondoh, in *Proceedings of IEEE International Symposium on Semiconductor Manufacturing (ISSM '03)*, p.207, San Jose, Calif, USA, (2003).
7. M. Ronay, *Journal of the Electrochemical Society*, p.G494, vol. 148 (2001).

Ex. 1016, Page 671

ECS Transactions, 18 (1) 645-649 (2009)
10.1149/1.3096515 © The Electrochemical Society

Etch Arts of Dual Damascus Structure and Their Impacts on WAT/VBD in 65nm Cu Interconnects

Wu Sun[a], Shih-Mou Chang[a], Hai-Yang Zhang[a], Xiaoming Yin[a], Liya Fu[a], Baodong Han[a], Xinpeng Wang[a] and Yongqin Wu[a]

[a] Logic Technology Development Center, Semiconductor Manufacturing International Corporation, No.18 Wen Chang Rd., BDA, Beijing, 100176, P.R.China

Abstract

In this paper, based on the via-first DD technology, we focus on the profile tuning of via and trench and the liner removal in order to optimize the trench profile. Via profile tuning include how to realize the via with the bowling profile, vertical profile and tapering profile. Trench profile tuning emphasizes the formation mechanism of bowling profile and top rounding profile. Footing profile and under-cutting profile of via bottom in liner open step is also addressed. The effect of different etching profiles on Rc, VBD (Breakdown Voltage) are examined. Results show that different via profiles have no impact on Rc, top-rounding trench-profile results in worse VBD, while under-cut & footing liner would deliver lower and higher Rc, respectively.

1. Introduction

As logic CMOS process continues to drastically scale down to 65nm node and beyond, Cu interconnects and low-k dielectrics have been widely integrated [1] in BEOL process instead of Al connection owing to the small conductance of Cu and the low dielectric constant of low k materials. Moreover, unlike direct Al etching technology, DD technology [2] has been extensively utilized owing to the nonvolatile characteristics of copper in dry etching.

In order to reduce the contact resistance （Rc） delay at the BEOL Cu interconnections, low-k materials, with dielectric constant lower than 3.0, has been widely used as the insulator in technologies for 90nm/65nm node and below. The porosity in low k insulators poses the new challenges in every aspect of the integration. During etching process, such low k dielectric tends to degrade when subject to conventional plasma processes such as etching and ashing in the fabrication of well-known dual damascus (DD) structure.

Back end of line (BEOL) DD process mainly consists of either via-first process or trench-first process. As the former features extendable patterning and easy control of via bottom critical dimension (CD), we focus on BEOL via-first process in this paper to clarify the effect of various etching processes on different via/trench profiles and the responses of corresponding Rc and VBD tests.

Ex. 1016, Page 673

2. Etching Experiments

Although hard-mask approach [3] could minimize the ashing damage, we still chose 65nm photo-resist (PR) mask approach for its easy integration to demonstrate the mechanism and method of etching profile tuning.

2.1 Effect of via etch on via profiles

As shown in Fig 1, via profile strongly depends on C4F8/O2 gas ratio. Either tapering profile or bowling profile can be seen if gas ratio of C4F8/O2 is not optimized. Higher ratio of C4F8/O2 is usually associated with heavier polymer generation. More polymer will be deposited on via sidewall, thus alleviating plasma bombardment and reducing sidewall etch rate. Finally tapering via profile is formed as shown in Fig 1 (left). In contrast, lower C4F8/O2 gas ratio results in polymer lean process. Via sidewall is subject to more bombardment. This enhances the etch rate of low-k sidewall and leads to more bowling via profile as shown in Fig 1 (right)..

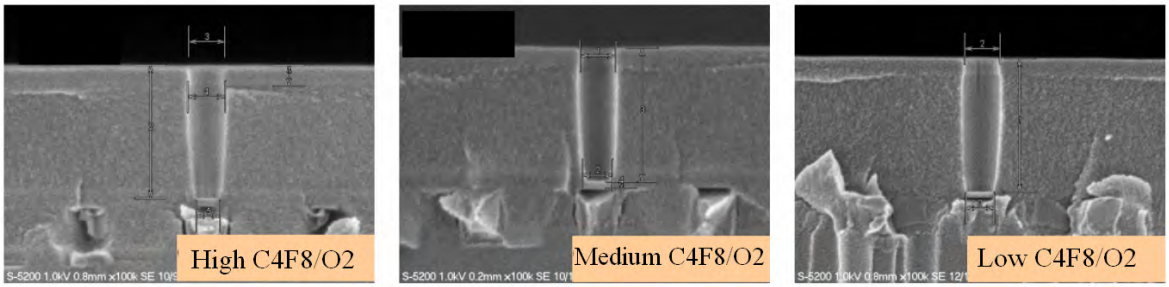


Fig.1 The effect of gas ratio on via profile

Fig 2 shows that compared with vertical and bowling profile, tapering via profile is associated with higher Rc. Uniformity of all the splits is comparable, though. The root cause can be ascribed to the difference of via bottom CDs owing to different via profiles. Straightforwardly, the tapering profile results in smaller via bottom CD among three splits with the same ADI CD. Smaller bottom CD will decrease the contact area of via bottom, thus delivering the higher Rc.

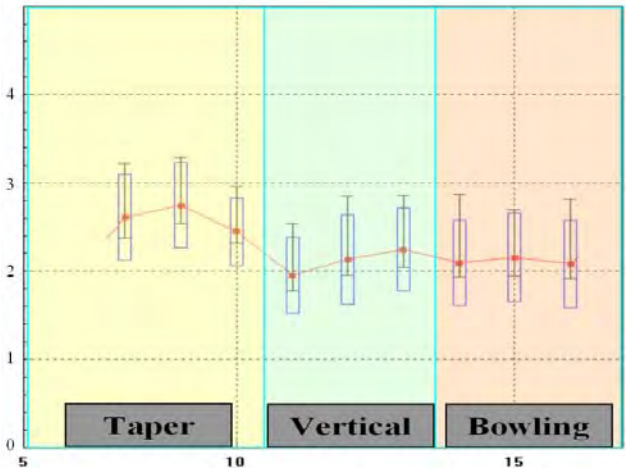


Fig.2 The effect of via profile on Rc

2.2 Effect of trench etch on trench profiles

The effect of polymer gas on trench profile is shown in Fig 3. Similarly, tapering trench profile can be fabricated if heavy polymer gas CH2F2 is utilized. While polymer lean process with CF4 gives more vertical trench profile. In fact, Cu deposition will benefit from vertical trench profile while polymer rich gas CH2F2 features worse non-uniform etching rate and is apt to result in worse trench depth uniformity.

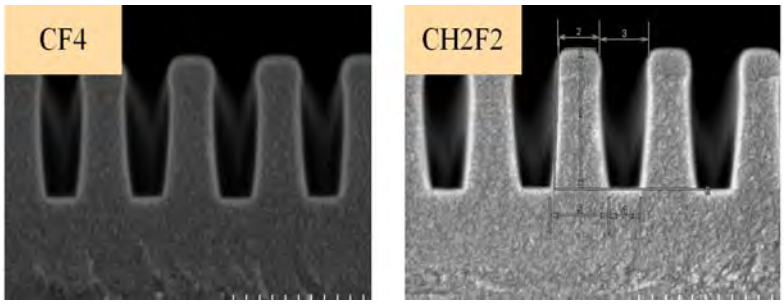


Fig.3 The effect of polymer gas on trench profile

Ashing process, a standard step to remove trench etch mask, is also a key knob for trench profile tuning. As shown in Fig 4, it shows that trench profile is vertical before mask is removed. However, slight bowling profile can be noticed on trench sidewall after PR mask is removed. It can be attributed to that huge amount of O2 in ashing process is used for traditional PR remove process. Low k material is easily damaged by O2 strip process, thus low-k sidewall is susceptible to O2 strip step and bowling profile appears.

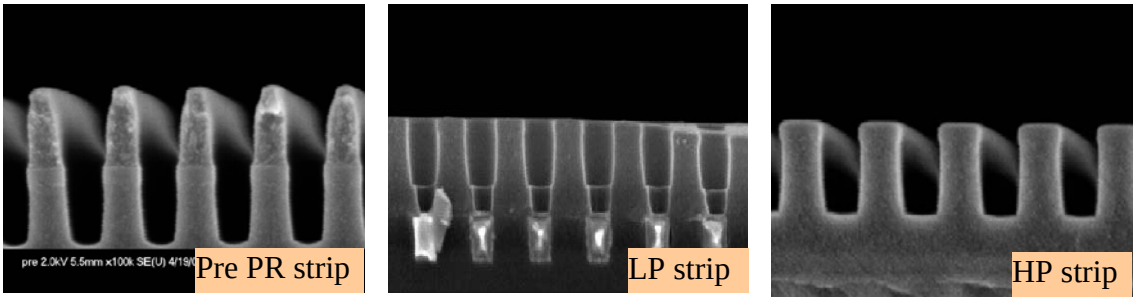


Fig.4 The effect of high and low pressure PR strip process on trench profile

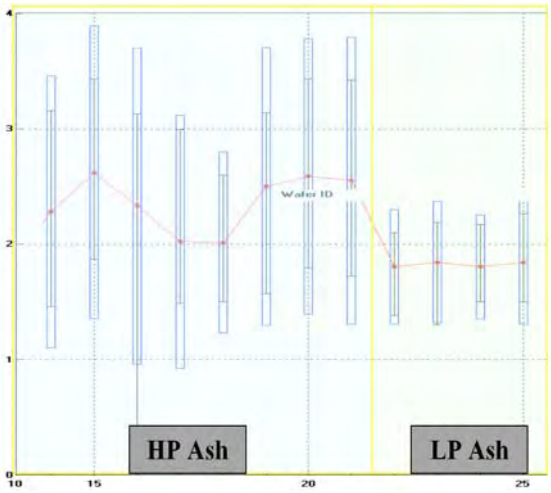


Fig.5 The effect of pressure of ash process on Rc

Fig.4 also demonstrates that high-pressure (HP) ashing process delivers more bowling profile than low-pressure (LP) ashing process. Their corresponding Rc performances are depicted in Fig 5. LP ashing shows more convergent Rc distribution than HP ash. Such phenomena can be attributed to the poor capability of via bottom cleaning from HP ash. While LP ash features more vertical bombardment, this enhances its tendency for more residue removal at via bottom. This finally leads to convergent Rc distribution.

2.3 Effect of stop layer remove on via/trench profiles

Fig.6 demonstrates that different etching gases in LRM step can lead to quite different trench top profiles and via bottom profiles. Smooth trench top profile and undercut profile at via bottom come from NF3 LRM process. This can be ascribed to the better selectivity (SiN to SiO2) of NF3 than CF4. Higher selectivity indicates less SiO2 at trench top will be consumed in NF3 process than CF4 process, therefore CF4 process tends to form a top rounding profile. Besides, NF3 LRM process is a polymer free process for no C element is available in NF3 gas. Its side effect is the weak sidewall protection at via bottom and is apt to form undercut as shown in Fig 6 (lower-left).

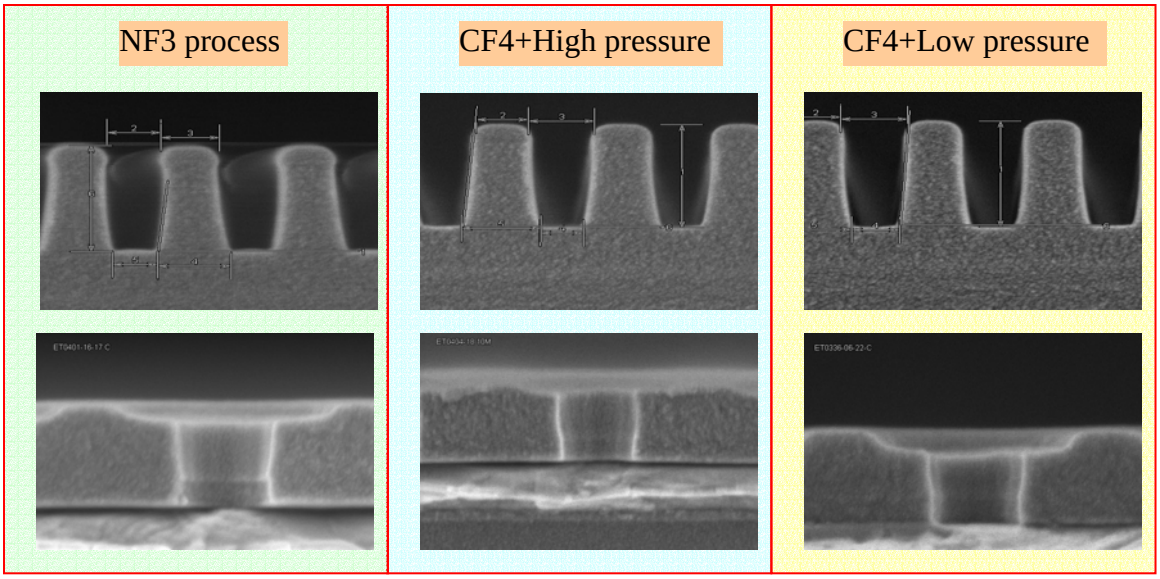


Fig.6 The effect of LRM process gas and pressure on trench and via bottom profile

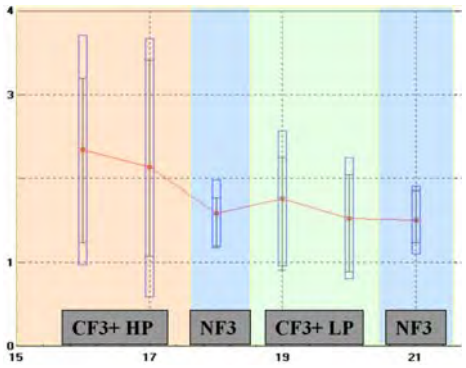


Fig.7 The effect of LRM process gas and pressure on Rc

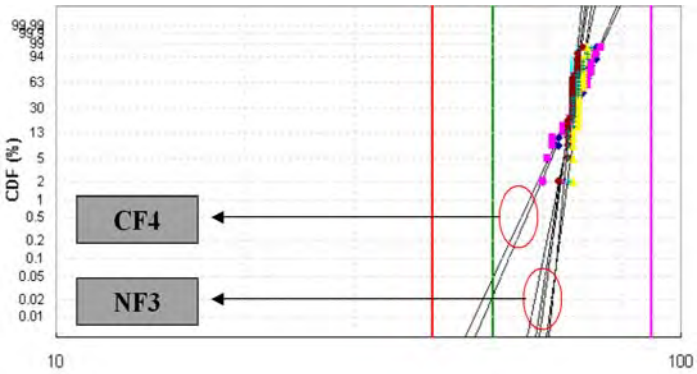


Fig.8 The effect of LRM process gas species on VBD

Normally, top rounding corresponds to the shrink of the top line-width CD. The weakest point of top rounding profile has been proven to result in worse VBD performance as shown in Fig. 8. Meanwhile, top rounding from polymer lean process might lead to worse sidewall roughness. All the above denotes LRM step is also a key point to well control VBD performance. Slighter sidewall roughness will deliver better VBD performance.

The best Rc performance comes from undercut profile (Fig.7) while footing profile gives the worst Rc performance. However, from the point view of reliability, neither undercut nor footing profile is the desired one. The biggest challenge is how to achieve the best Rc performance without either footing or undercut profile. Etching gas selection from the point view of lean/rich polymer and its uniformity will be the key to this target.

Conclusions

The effects of etching gases on via/trench profiles are evaluated based on the via-first DD technology. Polymer-rich etching gas has been proven to result in tapering profile for either via profile or trench profile. If CD and depth can be well controlled, tapering profile should be acceptable. Bowling trench profile usually comes from ashing process. Nevertheless, low pressure ashing process can alleviate this issue to some extent. CF4 LRM process can result in top rounding profile for its low selectivity of SiN to SiO2.Top-rounding profile is apt to cause worse VBD. Polymer-lean NF3 LRM process tends to bring undercut trench profile. This also needs avoid for the sake of reliability concern.

Acknowledgments

The authors wish to thank etch department of logic technology development center of S.M.I.C. for some additional data and share of their theory and experience.

References

1. Y. Hayashi, et al, Proc. Of IITC, 2002, p145 (2002)
2. M. Tada, et al., IEDM Tech. Dig, p845 (2003)
3. O. Hinsinger, et al., IEDM Tech. Dig, p317, (2004)
4. Wu Sun, et al., Mechanism of Via Etch Striation. ICSICT2008. (2008)

ECS Transactions, 18 (1) 651-658 (2009)
10.1149/1.3096516 © The Electrochemical Society

Understanding and Importance of Defects in Advanced Materials

Luigi Pantisano, [b] L.Trojman, [b] M.Aoulaiche, [c] R.O'Connor, B.Kaczer, [b] G.Groeseneken
IMEC Kapeldreef 75, Leuven (Belgium),
[b] also Electrical Eng. Dept., KULeuven, Leuven (Belgium)
[c] Dublin City University, Dublin (Ireland)

Aggressive CMOS scaling and device performance enhancement has been pursued through the use of advanced dielectric and substrate materials. High-k dielectric materials are the key to both dramatically reduce the standby leakage current (and thus power consumption) and to enable the use of high mobility substrates (Ge, III-V, etc). Critical for obtaining high device performance and satisfactory reliability are the interfacial quality and the bulk defect control. In this paper, electrical characterization techniques, analyzing the nature and impact of defects on device performance and reliability, are presented.

Introduction

The introduction of new materials in semiconductor processing leads to certain trade-offs, such as new defects. Among these defects, electrically active defects are one of the key concerns and need to be understood and controlled. Therefore the corresponding processes and defect control require advanced characterization and modeling. In particular for nanotechnology, interface and near-interface defects play a key role in controlling degradation and wear out mechanisms. In this contribution we show three showcases for defects in advanced high-k gate stack materials, focusing specifically on near-interface defects since they affect the device performance as well as the reliability.

Samples considered in this paper feature high-k / metal gate materials are deposited on two different Si-orientations: Si-(110) and Si-(100). Note that Si-(110) has a higher density of bonds at the interface thus resulting in a faster dielectric nucleation during deposition as well as high interface state density. By using a combination of mobility techniques ([1] and references therein) and a low frequency charge pumping method [2], the near interface dielectric quality can be reliably studied. These advanced techniques can also be applied to Germanium devices [3].

Near interface and bulk defects also play a significant role in reliability phenomena. Bias temperature instability (BTI) and wear out phenomena limit the lifetime of these devices. During a BTI test, the defects are charged and discharged dynamically with the applied bias. This indicates that these near interface defects, closely associated with the chemical composition at the interface (namely N and Hf concentrations), should be correctly accounted for in the BTI lifetime extrapolation. In the third and the last example, the reliability implications of defects are studied using Stress Induced Leakage current (SILC) spectroscopy. SILC shows both where these defects are located (energy and spatial location), which are crucial in the understanding of the wear out phenomena.

Experimental

Identical CMOS devices were prepared on 200mm wafers on either (110) or (100) Si-surface orientation. The current flow direction was <110> for all MOSFETs. The HfSiON gate stacks were formed by an initial IMEC clean, which results in a 1nm chemical oxide, followed by an ALD or MOCVD HfSiON growth. Several nitradation schemes were used, being either a thermal or by a DPN. As gate electrode, either TiN or TaN gates were used and then capped with poly-Si for electrical contact.

Interface Quality for 1nm EOT HfSiON on Si-(110)

PFET on Si-(110) have recently attracted a lot of attention due to the high mobility – comparable to the one of NFET on (100) [4], as shown in Fig. 1. Si-(110) has 3x higher bond density thus resulting in a higher interface states density and a faster nucleation of high-k. Therefore it's not obvious whether having the same mobility for PFET on Si-(110) and NFET on Si-(100) can be considered as an advantage from the reliability perspective since the N_{it} and the corresponding defects in high-k may be radically different. This is a particular concern for scaled devices in the 1nm-EOT range where the interfaces dominate the overall dielectric contribution.

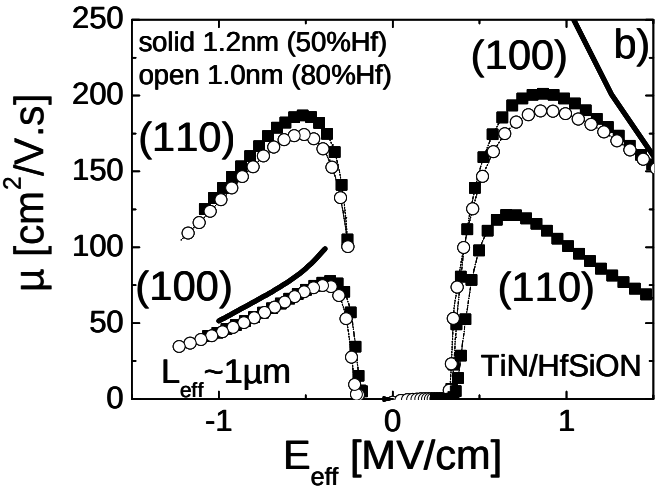


Fig. 1 Si-(110) has a larger (smaller) hole (electron) mobility compared to the same devices on Si-(100).

Two experiments were designed to quantify the near interface defects. In the first experiment (Fig. 2) the mobility has been measured in a regime which is sensitive to remote coulomb scattering (RCS) and phonon scattering. By lowering the lattice temperature, the RCS remains while the phonon contribution for both Si-(110) and Si-(100) is significantly reduced. Note that both NFET-(100) and PFET-(110) have the same temperature-dependent mobility increase, which strongly suggests that they share very similar scattering mechanisms (i.e., defects). In the second experiment, a charge pumping method was used to assess interface and near-interface defects [2], as shown in Fig. 3. Note a higher N_{it} for Si-(110) is observed, as expected due to the corresponding higher bond density. From the results of Figs. 2-3, we can safely conclude that the near-interface high-k quality is maintained by changing from the Si-(100) to a new orientation of (110).

Results in Figs. 1-3 demonstrate the importance of combining together different approaches to determine dielectric qualities. Nevertheless these results are rather surprising since one may wonder why semiconductors where high-k nucleates faster (i.e., less interface defects) feature the same interfacial quality as a standard Si. After analyzing in more details the EOT and device characteristics (see [1]), we found that the EOT was unchanged. This indicates that the interface reorganization during the thermal annealing (i.e., after high-k deposition). In addition, we also found out that the choice of the starting interface growth dominates the overall electrical response. Therefore the interface reconstruction after thermal annealing is the key parameter which should be closely monitored.

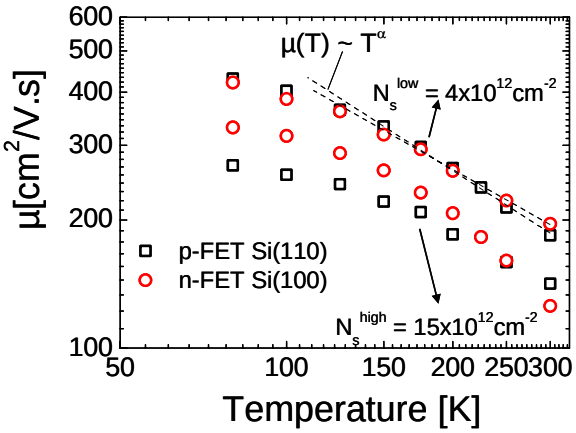


Fig. 2 Temperature dependence of mobility. The mobility is measured in a region where high-k defects impact is the highest (i.e., Remote coulomb scattering – RCS component). The same temperature dependence of RCS for nFET-(100) and pFET-(110) indicates that the high-k dielectric quality is the same.

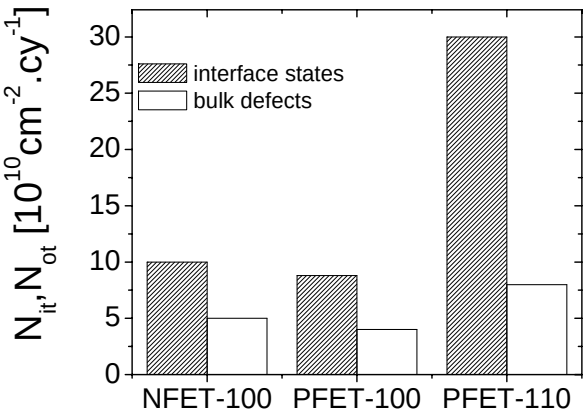


Fig. 3 The Charge Pumping method is able to extract the interface states N_{it} and the bulk defects. Note that p-MOS on (110) shows 3 times larger interface state density.

Defects in the Near Interface – Impact on BTI

As shown in Fig.1-3, mobility, an important parameter representing device quality, is observed to be reduced when nitrogen is incorporated into the device. This is a typical case where near surface defects become important and cause instability. When pFETs are biased high (i.e., inversion) at high temperatures, a Bias Temperature Instability (BTI) is observed since the corresponding threshold voltage (VT) shifts during time. When the bias is removed (Fig. 4), some of the initial VT is recovered. The measured relaxation depends very much on the time resolution of the measurement setup [5, 6].

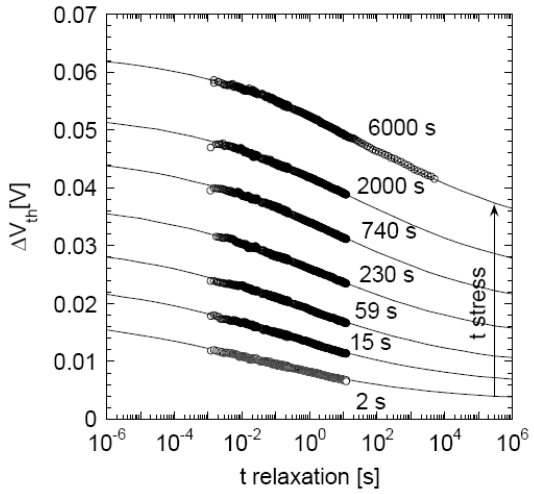


Fig. 4 ΔV_{th} as a function of the relaxation time for different stress times. The continuous lines are obtained from the fitting using the universal recovery behavior [6]

Several attempts have been made to model and quantify the transient effects and the component that remains permanent. Note that the permanent component has been associated with a longer term reliability issue (i.e., irreversible damages). Conversely the recoverable component is associated more with the near-interface defects where they can be charged / discharged during the measurement and are associated with the gate stack quality (i.e., chemical composition, processing, and impurities). This is indeed for the case of Fig. 5 where the same HfSiO is compared with or without nitrogen incorporation. While N is beneficial for lowering leakage and EOT reduction, it may have a negative effect on the BTI. Nitrogen incorporation in the near interface seems to generate more defects which could be charged / discharged.

To gain insight into the defects responsible of the NBTI degradation kinetics, the charge pumping current (CP) and stress induced leakage current (SILC) spectroscopy were studied. In SILC spectroscopy, a large bias is intentionally applied to generate defects in the dielectrics, which leads to an increase in the gate current (Fig. 6). Following the striking similarities with similar experiments done with SiO_2, we found that such an increase is due to oxide defects which act as a stepping-stone thus leading to a trap-assisted current increase. The SILC spectroscopy of Fig. 6 indicates that the N incorporation introduces a defect with the energy level within the silicon mid gap, which is consistent with the K_N center reported by J.P Campbell et al [7].

Ex. 1016, Page 682

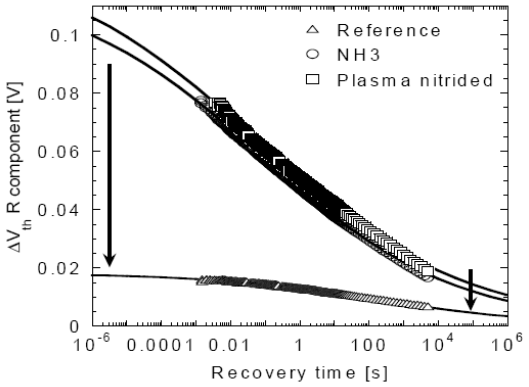


Fig. 5 The remaining R component of the ΔVth as a function of recovery time shows a higher Recovery component in the nitrided stacks. The continuous lines are the fit using the universal recovery behavior [8]

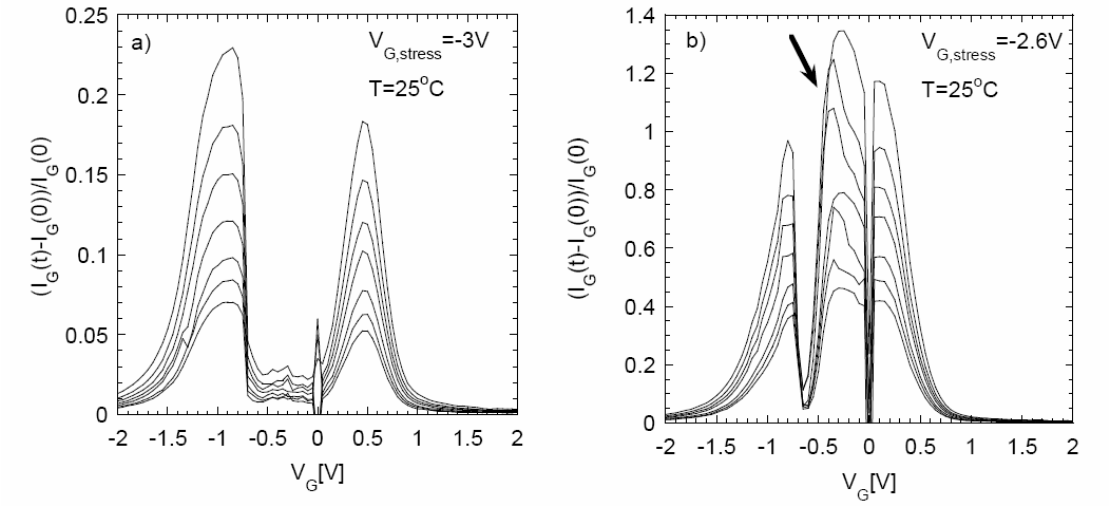


Fig. 6 a) SILC spectrum of the reference sample stressed with increasing stress time and b) of one of the nitrided stacks. An additional peak in the SILC is appeared from the nitrided stack (i.e., comparing the left to the right panel).

Defect Generation and Its Link to Device Wear Out

In this section, we plan to identify which defects are responsible for wear out and breakdown. Conventionally, a large bias is applied to the gate until the device breaks down. In addition, these devices are biased in a way so that the Fowler-Nordheim regime is reached. In the F-N regime, the injection fluence is also controlled by the electric field at the injecting interface. This implies that the applied bias controls both the fluence and the injected carrier energy. Therefore it's difficult to decouple whether the defects are created by the effect of the fluence or by the injected electrons energy.

SILC spectroscopy can also be used to understand these reliability problems for high-k films. Results similar to those of Fig. 6 revealed that electron traps are created close to the Si conduction band. However, SILC spectroscopy alone does not distinguish very

Ex. 1016, Page 683

well the defect spatial location since the electrical signal is dominated by the leakiest traps.

Substrate hot electron injection (SHE) can be used to experimentally modulate the spatial location where traps are generated (i.e., interface vs. bulk HfSiON), and therefore their nature. In a nutshell, SHE consists of accelerating substrate electrons with a large bias (see Fig. 7(a)) toward the gate contact, while the MOSFET is biased in inversion ([9, 10] and references therein). Using SHE has two major advantages: during stress the electric field can be independently controlled and the hot electron energy can be varied independently from the fluence. If the substrate bias provides adequate energy to the electrons (i.e., a potential larger than the Si/SiO_2 barrier height), some electrons are injected over the barrier and thus create defects in the dielectric. This is the case shown in Fig. 7(b) where at low electron energy N_{it} are preferentially created while at high energy bulk defects are created.

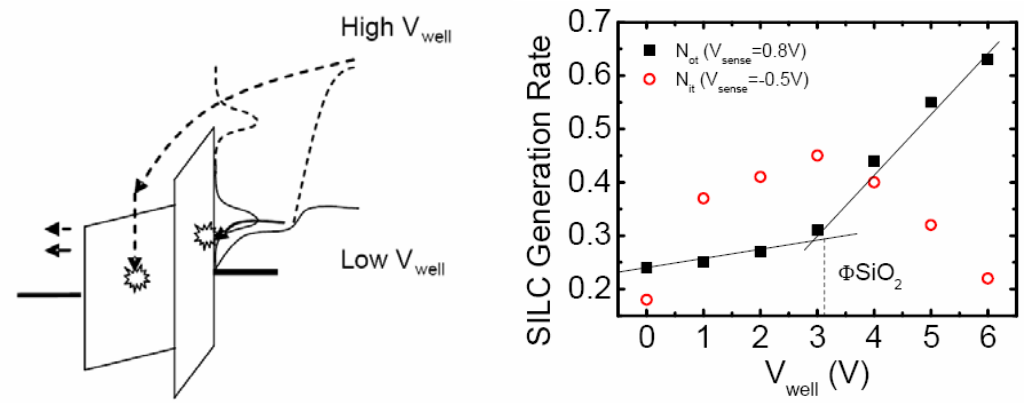


Fig. 7 a) A schematic illustration of the substrate hot electron (SHE) injection technique, in which a low Vwell primarily damages the interface while a high Vwell causes bulk damages. b) Corresponding interface and bulk trap generation rates extracted from the SILC spectrum are consistent with the dominant mechanism during the respective substrate hot electron stresses

SILC spectroscopy revealed two kinds of defects created during stress. One may wonder which one is responsible for the occurrence of breakdown (see Fig. 8). In Fig. 8, the breakdown data were collected as a function of different SHE injection conditions, namely high and low energy and then compared with the standard constant voltage stress (CVS). The matching between the low SHE energy and CVS data suggests a similarity in the degradation mechanisms. Note in Fig. 8 the dramatic acceleration of the charge to breakdown as well as the change in Weibull slope compared to standard constant voltage stress. The Weibull slope, a measure of the bulk trap generation, is proportional to the defect generation rate here. Results of Fig. 8 suggest that high-energy electrons are more damaging. This is indeed the case when the defect generation mechanisms, correlated with the breakdown mechanisms as shown in Fig. 9, has demonstrated that bulk defects (as measured by SILC) are the ones also responsible for breakdown.

Ex. 1016, Page 684

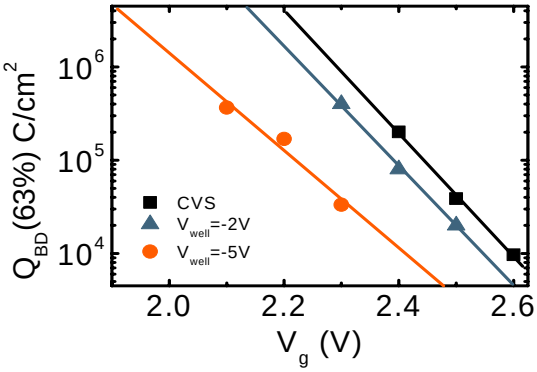


Fig. 8 Charge to breakdown Q_{BD} and gate voltage acceleration for a 710^{-8}cm² device are reduced significantly for stresses with high hot carrier energy (V_{well}), consistent with the behavior of the SILC peak found at 0.3eV below E_c^{Si}, suggesting these traps dominate the breakdown.

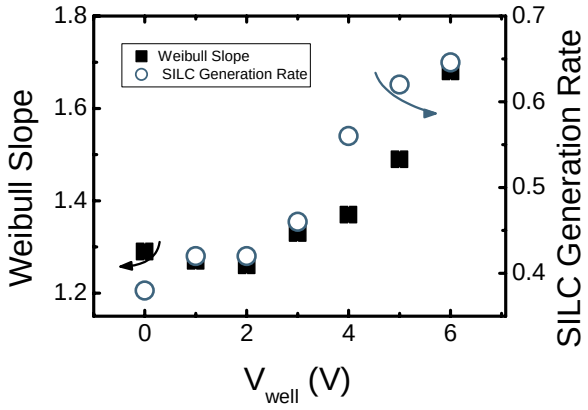


Fig. 9 Comparison of Weibull slope and the generation rate of the SILC peak attributed to bulk traps during substrate hot electron stresses at a range of well voltages. The Weibull slope is proportional to the bylk states responsible of the breakdown and it's mirrored by the SILC behaviour (at the SILC bias where bulk states are correctly measured).

Conclusions

Quantifying and understanding the nature of defects and their impacts on process technology continue to remain an important topic for future research and technology development. The experimental results shown in this paper validate the effectiveness of combining different techniques and approaches (reliability, electrical and physical analysis, etc) toward the understanding of interfacial and near surface defects of semiconductor/hi-k oxides.

Acknowledgements

This work was carried out as part of IMEC's Industrial Affiliation Program, funded by IMEC's core partners: Intel, Texas Instruments, Micron, Infineon, Qimonda, NXP, ST, Matsushita, TSMC and Samsung. Robert O'Connor wishes to acknowledge the EU Marie Curie Intra European Fellowship Scheme for funding.

References

1. L.Trojman, L.Pantisano, I.Ferain, S.Severi, H.Maes, and G.Groeseneken, IEEE Trans. El. Dev. vol.55, no.12, pp.3414-3420, Dec. 2008.
2. M.B. Zahid, R. Degraeve, L. Pantisano, J.F. Zhang , and G. Groeseneken, in *Reliability physics symposium, 2007, IEEE proceedings.*, vol. 45, pp.55-60.
3. J. Mitard, B. De Jaeger, F.E. Leys, G. Hellings, K. Martens, G. Eneman,D.P. Brunco, R. Loo, J.C. Lin, D. Shamiryan, T. Vandeweyer, G. Winderickx, E. Vrancken, C.H.Yu, K. De Meyer M. Caymax, L. Pantisano, M. Meuris, M.M. Heyns, IEEE International Electron Devices Meeting, San Francisco, CA, USA, 2008.
4. B. (Frank) Yang, A. Waite, H. Yin, J. Yu, L. Black, D. Chidambarrao, A. Domenicucci, X. Wang, S. H. Ku, Y. Wang, H. V. Meer, B. Kim, H. Nayfeh, S.D. Kim, K. Tabakman, R. Pal, K. Nummy, B. Greene, P. Fisher, J. Liu, Q. Liang, J. Holt, S. Narasimha, Z. Luo, H. Utomo, X. Chen, D. Park, C.Y. Sung, R. Wachnik, G. Freeman, D. Schepis, E. Maciejewski, M. Khare, E. Leobandung, S. Luning, P. Agnello, IEEE International Electron Devices Meeting, Washington DC, USA, pp. 151-154, 2007.
5. V. Huard, C. Parthasarathy, N. Rallet, C. Guerin, M. Mammase, D. Barge, C. Ouvrard, IEEE International Electron Devices Meeting, Washington DC, USA, 2007.
6. B. Kaczer, T. Grasser, Ph. J. Roussel, J. Martin-Martinez, R. O'Connor, B. J. O'Sullivan, G. Groeseneken, *Reliability Physics Symposium, 2008, IEEE proceedings.*, vol. 46, pp.28-38, April 27 2008-May 1 2008.
7. J.P Campbell and P.Lenahan, *Reliability Physics Symposium, 2007. IEEE proceedings.*, vol. 45, 503-509.
8. M.Aoulaiche, B. Kaczer, Ph. J. Roussel, and R. O'Connor, M.Houssa, S. de Gendt, H. E. Maes and G. Groeseneken, J. Vac. Sci. Technol. B **27 (1)**, 2009 1071-1023.
9. R.O'Connor, L.Pantisano, R.Degraeve, T.Kauerauf, B.Kaczer, P.J.Roussel, and G.Groeseneken, J. Appl. Phys. 103, 064503 (2008).
10. R. O'Connor, L. Pantisano, R. Degraeve, T. Kauerauf, B. Kaczer, Ph. J. Roussel, G. Groeseneken, *Reliability Physics Symposium, 2008, IEEE proceedings.*, vol. 46, 324-329.

ECS Transactions, 18 (1) 659-662 (2009)
10.1149/1.3096517 © The Electrochemical Society

Yield Enhancement with Optimized Offset Spacer Etch for 65nm Logic Low-leakage Process

Bao-Dong Han[a], Shihmou Chang[a], Hai-Yang Zhang[a], Lin-Lin Zhao[a], Ya-Li Fu[a], Wu Sun[a], and Bing-Wu Liu[a]

[a] Semiconductor Manufacturing International (Beijing) Corp (SMIC).
No. 18 WenChang Road, Beijing, 100176, P.R. China

Abstract

This paper presents an effective way to reduce the degradation of the Ring-OSC speed with an optimized offset spacer etch, which renders an improvement of the within-wafer uniformity of the offset spacer width and a reduction of the corresponding profile deviation. We have developed a lower offset spacer etch rate scheme for better control of plasma uniformity. In comparison with the 3σ of 1.3nm offset spacer width uniformity and 1.8nm profile deviation of the conventional dielectric etch with high etch rate, this new low etch rate process has delivered a 3σ of 0.5nm for the offset spacer width uniformity and 0.7nm offset spacer profile deviation. The corresponding Ring-OSC loss has been completely eliminated and the wafer-level yield is enhanced by 40%.

1. Introduction

During the process development of 65nm logic technology, the low-leakage performance without compromising the transistor speed is required. No degradation of transistor speed as the trade-off is allowed in meeting the requirement of the low leakage device performance. Transistor speed loss could be monitored using the Ring Oscillator which is strongly correlated to the gate-to-drain overlap capacitance (CGD0). In order to reduce the scaling impacts to junctions and alleviate the overlap capacitance without at the expense of the junction resistance [1-2], offset spacer is formed at the sidewall of the gate, and it has been widely-used since 90nm CMOS nodes. The offset spacer and S/D extensions should be carefully optimized in a way to avoid either the loss of the drive current or short channel effects. To obtain the acceptable repeatability and parameter dispersion in an offset spacer based CMOS process, the uniformity of offset spacer width must be controlled to less than 1nm [2]. This stringent control limit has been tightened further for 65nm node and beyond processes [3]. Fig.1 shows that Ring-OSC loss is strongly correlated to the CGD0 shift and the latter is related to offset spacer width. The improvement of offset spacer width uniformity and the deduction of its profile deviation can reduce within-wafer CGD0 variation and hence the Ring-OSC loss. In this paper, we present the uniformity and profile deviation improvement results of 65nm offset spacer etch from a newly optimized process using a silicon etcher. The Ring-OSC loss has been significantly reduced and the corresponding wafer-level yield is improved by 40%.

Ex. 1016, Page 687

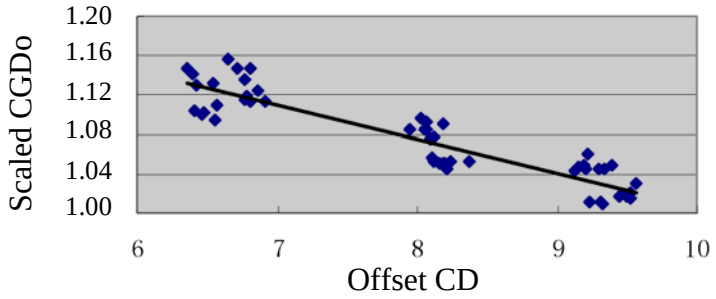


Fig.1 NMOS CGD0 VS OFFSET CD

2. Experimental results

A schematic of an offset spacer film stack is illustrated in Fig.2. The offset spacer etch process consists of a main etch and an over etch step.

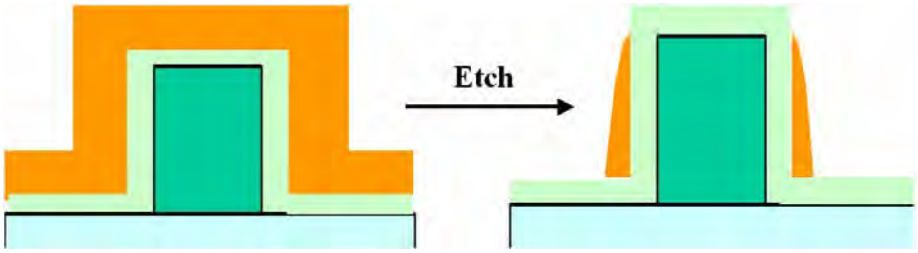


Fig.2 Offset spacer film stack

The offset etch process tuning was carried out using a conventional dielectric etcher and a Lam 2300 Versys Kiyo. Conventional dielectric always gave a high etch rate, while Lam 2300 can provide low etch rate process. Kiyo also features additional knobs, such as wafer temperature, gas injection, and a symmetric conductance. The optimization using these knobs provides the low energy bombardment and tunable center/edge process. A comparison of these two tools is shown in Fig.3 and Table 1. Fig.3 shows that the offset spacer width uniformity has been greatly improved to a tight 3σ distribution of 0.5nm from Lam 2300 Versys Kiyo, while the conventional dielectric etcher, with a higher etch rate, could not achieve a 3σ value better than 1.3nm. In addition, the offset profile deviation (defined as bottom CD – middle CD) has been reduced to 0.7nm from Lam 2300 Versys Kiyo while a large value of 1.8nm was obtained from the conventional dielectric etcher (Table 1).



Fig.3 Offset spacer width uniformity comparison

Table 1 Offset profile deviation comparison

	CDbottom-CDmiddle(nm)	
Wafer	Lam 2300 Versys Kiyo	Conventional etch tool
1	0.9	2
2	0.5	2.5
3	0.3	1.7
4	1	2.4
5	1	2.2
6	0.6	0.3
Average	0.7	1.8

Ex. 1016, Page 688

Ring-OSC loss results in Fig.4 indicate that the improved offset spacer width uniformity significantly reduces ring-OSC loss. The corresponding wafer-level yield data show also a significantly improvement of 40%. The Ring-OSC CDF curve in Fig.5 also demonstrates offset etch with low etch rate delivers a closer-to-target performance. The Universal curve and V_t Roll-off curve show that the low etch rate process from Lam 2300 Versys Kiyo outperforms the high etch rate of conventional dielectric tool on N-MOS, while little difference on P-MOS, as shown in Fig.6. This is due to the faster boron diffusion in P-MOS rendering them insensitive to offset spacer width uniformity.

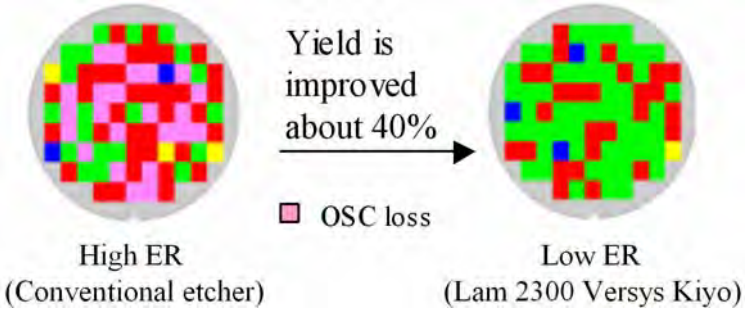


Fig.4 Ring-OSC loss and yield results comparison

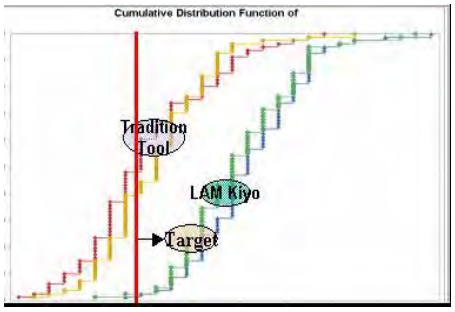


Fig.5 Offset spacer etch split OSC CDF

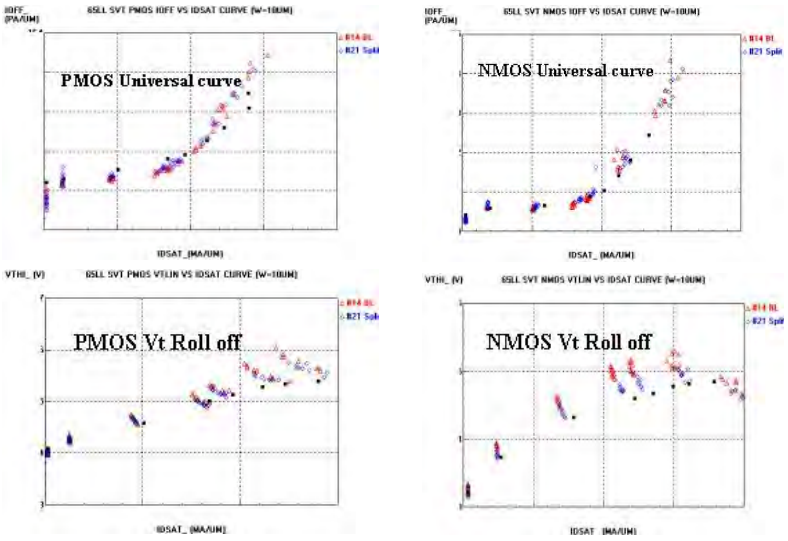


Fig.6 PMOS and NMOS performance comparison

Idsat is another yield loss factor related to the offset spacer width uniformity. Fig. 7 shows that a 1nm offset spacer width change can result in a ~5% Idsat shift on N-MOS. It is evident that yield enhancement will benefit from a tight 3σ offset spacer width.

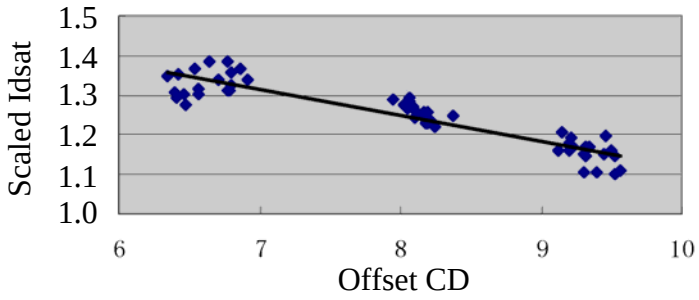


Fig.7 NMOS IDSAT VS OFFSET CD

3. Conclusions

Significant yield enhancement has been demonstrated for 65nm low-leakage device using a low etch rate process developed on a Lam 2300 Versys Kiyo. The resulted improvement on width uniformity and profile deviation of offset spacer etch reduces greatly the Ring OSC loss and Idsat shift, which leads to a yield enhancement of 40%.

Acknowledgments

Authors would like to thank Chun Wang and Song Huang of Lam Research China for the technical discussion and information.

References

1. Goto, K. Tagawa, Y. Ohta, H. Morioka, H. Pidin, S. Momiyama, Y. Kokura, H. Inagaki, S. Tamura, N. Hori, M. Mori, T. Kase, M. Hashimoto, K. Kojima, M. Sugii, T. "High performance 25 nm gate CMOSFETs for 65 nm node high speed MPUs" IEDM Tech Dig., pp. 623-626, (2003).
2. Hokazono, A. Ohuchi, K. Takayanagi, M. Watanabe, Y. Magoshi, S. Kato, Y. Shimizu, T. Mori, S. Oguma, H. Sasaki, T. Yoshimura, H. Miyano, K. Yasutake, N. Suto, H. Adachi, K. Fukui, H. Watanabe, T. Tamaoki, N. Toyoshima, Y. Ishiuchi, H. "14 nm gate length CMOSFETs utilizing low thermal budget process with poly-SiGe and Ni salicide", IEDM Tech Dig., pp. 639-642, (2002).
3. Jurczak, M. Rooyackers, R. De Keersgieter, A. Kunnen, E. Henson, K. Richard, O. Dachs, C. "Triple junctions for reduced impact of offset spacer variation on CMOS device parameters", ESSDERC 2004, pp.145-148, (2004).

Ex. 1016, Page 690

ECS Transactions, 18 (1) 663-668 (2009)
10.1149/1.3096518 © The Electrochemical Society

STUDY OF FILM GROWTH ON SILICON IN ADVANCED HIGH DOSE IMPLANT PHOTORESIST STRIP

Li Diao and Songlin Xu

Mattson Technology, Inc, Fremont, CA 94538, USA

Downstream high dose implant photoresist strip (HDIS) plasma processes tend to cause an oxide growth on the exposed poly Si and a poly-Si loss, both need to be carefully controlled/minimized to avoid undesirable impacts to the corresponding IC device performance. Ellipsometry and FTIR spectrometry were used to characterize the ultrathin film growth and poly-Si loss introduced by the various chemistries used in HDIS, including O_2, O_2/FG, O_2/H_2, and N_2/H_2. Some discrepancies were observed between results obtained from Ellipsometry and those from FTIR for ultra thin films. Oxide growth was evident in O_2 dominant chemistries, accompanied by a low level poly-Si loss. No oxide growth was observed in non-O_2 chemistries. Hydrogen addition provided a better clean and a shorter processing time, but at the cost of increasing poly-Si loss and oxide growth.

Introduction

High dose implant photoresist strip (HDIS) in plasma reactors has been used extensively to remove photoresist (PR) in the front-end-of-line (FEOL) of device fabrication. High dose ion implantation introduces a tough crust on top of the PR, a highly cross-linked carbonized material mixed with implanted species and sputtered silicon/silicon oxide. It has become necessary to use not only gas mixture of both oxidizing and reducing chemistries instead of just oxygen, but also multiple process steps to remove this crust completely. The corresponding complicated chemistry and process recipes render the process tuning difficult, especially in controlling simultaneously the loss of silicon substrate and the growth of oxide films, and increases the device performance variation [1]. The International Technology Roadmap for Semiconductor (ITRS) calls for a mere few angstroms of silicon loss for 45nm node and beyond [2,3]. In order to meet this stringent goal, we have carried out an extensive study on the film growth and poly-Si loss during HDIS in various chemistries and process conditions. The corresponding film thickness and composition changes are monitored by Ellipsometry and Fourier transform infrared (FTIR).

Experiment

This work was performed in a SUPREMA downstream plasma tool from Mattson Technology, as shown in Fig. 1. Downstream plasma with reactive free radicals is obtained in the lower portion where ionic species are quenched by the grid. The test sample sits on a pedestal and its temperature is controlled by the pedestal temperature set by a heater.

Ex. 1016, Page 691

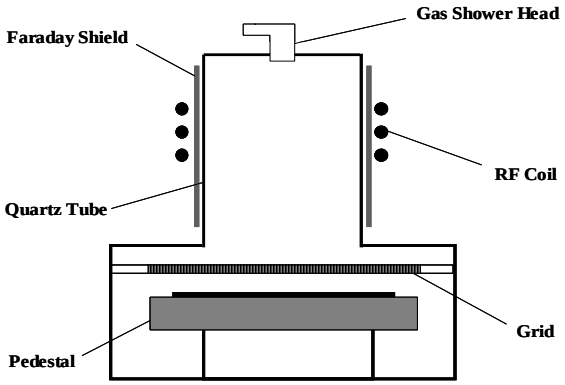


Fig. 1 Schematic diagram of a SUPREMA downstream plasma reactor

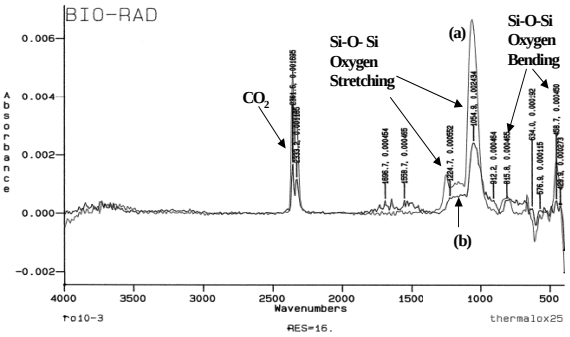


Fig. 2 FTIR spectra of (a) RTP oxidation (25Å) and (b) after O_2 plasma exposure at 275°C for 120s

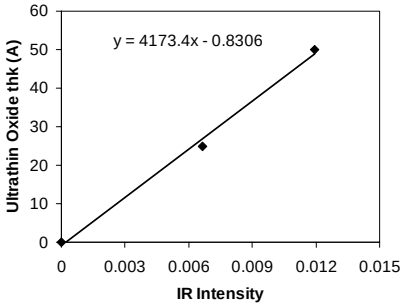


Fig. 3 Relationship between ultrathin oxide thickness and IR intensity

A FTIR transparent p-type (100) single crystalline Si wafer was used for film growth test, which was evaluated using an ultraviolet ellipsometer (Nanometrics 8300X) and a FTIR spectrometer (BIO-RAD QS-408M). Oxide thickness was measured before and after the plasma exposure. FTIR spectrum was obtained by a pre and post scan on the same sample and film property was monitored by the corresponding FTIR absorption (Fig. 2). Thermal oxide film was prepared using a rapid thermal process (RTP) with standard thermal oxidation at 900°C, which yields two thicknesses, namely 25Å (1065.6 cm^{-1}/0.006688, 1251.5 cm^{-1}/0.001356) and 50Å (1068.8 cm^{-1}/0.01193, 1254.1 cm^{-1}/0.00221). IR absorptions are in agreement with previous studies and similar red-shift of the absorption was also observed for both thin films [4]. A linear relationship, presented in Figure 3, was found between film thickness and IR intensities at 1050-1070 cm^{-1}. This suggests that the shift of the IR peak can now be used for estimating the ultrathin oxide thickness and corresponding absorption intensity can also be used to quantify the absolute film thickness. Poly silicon sample used for this work has a thick poly silicon (~1000A) blanketed on top of a ~1000A thermal oxide. The poly-Si silicon thicknesses were measured before and after is oxidation tests using Nanometrics and the difference obtained is therefore equal to the poly-Si loss. All the samples were treated with HF dip in a 100:1 DI:HF solution before stored in a N_2 box until usage. Unless stated otherwise,

the plasma process was carried out in an environment of 800mTorr with a RF power of 2500W and a pedestal temperature of 275 °C for 120s.

Results and Discussion

Chemistry Effect

Oxide growth is evident with O_2 containing chemistries, as shown by Fig. 4(a). In addition, it shows that the addition of forming gas (FG, 4% H_2/N_2) or H_2 enhances the oxide growth, especially for 20% H_2 which has more than doubled than that of pure O_2 per ellipsometer measurement. It is worth noting that the oxide growth from the non-O_2 plasma exposure was observed by the ellipsometer measurements, but not from FTIR spectra of Fig. 4(b). FTIR results revealed that only samples exposed to O_2 containing plasma had distinctive oxide absorption signatures at around 1050-1070 cm^{-1}, but no absorptions for either SiO_x or $Si_xO_yN_z$ like films were found after H_2, N_2 or H_2/N_2 mixture plasma exposure [5,6]. The above observation prevented us from using the ellipsometer for film growth measurement of non-O_2 chemistries. Poly-Si measurements showed no poly-Si loss with N_2 plasma exposure, as expected. A small ~4-5Å of poly-Si loss was obtained using O_2 chemistry with a minute addition of FG, which is a commonly used chemistry for HDIS process. However, significant increase of poly-Si loss with H_2-containing plasma exposure was observed no matter whether O_2 or N_2 was used. Among them, H_2/O_2 mixture doubled the poly-Si consumption compare to pure O_2 plasma exposure.

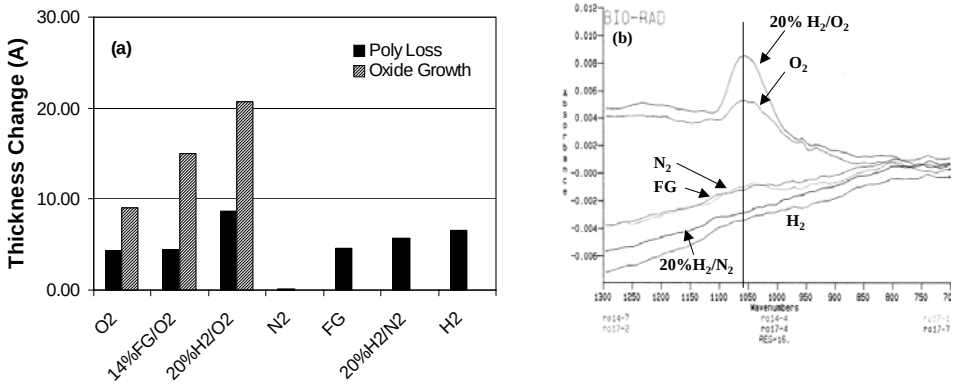


Fig. 4 Oxide growth and poly-Si loss from (a) ellipsometer and (b) FTIR spectra after the 120s plasma exposure at 275°C with different chemistries.

O_2 Chemistry

Figure 5 shows that the oxide growth increases linearly with the downstream O_2 plasma exposure time from the ellipsometer measurements. However, FTIR intensity results shows that the growth is exponential with a large thickness leap from 60 to 120s. We suspect that accuracy of oxide thickness less than 6Å is less reliable for measurements from ellipsometer. Poly-Si loss follows the linear trend up to 4.5Å. An increase of process pressure resulted in slight increase of oxide growth from ellipsometer and FTIR measurements (Fig. 6), whereas poly-Si loss seemed to be saturated at about 800mT. RF

Ex. 1016, Page 693

power had limited effect on oxide growth and poly-Si loss, as shown in Fig. 7. Pedestal temperature did not seem to affect oxide growth and poly-Si loss below 275°C, as shown in Fig. 8(a). Significant oxide growth was observed at 325°C with the ellipsometer. However, IR absorptions from Fig. 8(b) show no difference between that of 275°C and 325°C. This discrepancy has been attributed to certain surface changes which interfere with the ellipsometer. Overall oxide growth does not exceed 10Å and poly-Si loss is no more than 5Å under 120s plasma exposure, while shorter exposure dramatically reduces the oxide growth and poly-Si loss. Clearly, process time would be the most effective knob to control oxide growth and poly-Si loss, when O_2 chemistry is used as part of HDIS process.

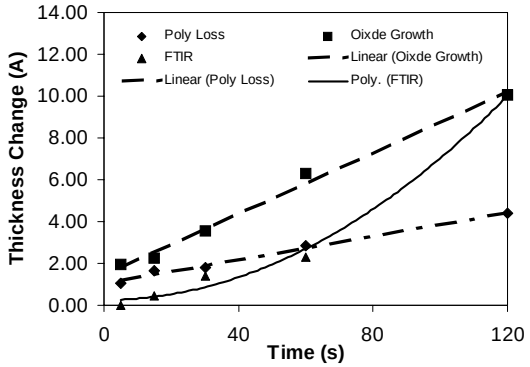


Fig. 5 Oxide growth and poly-Si loss as a function of exposure time in O_2 plasma at 275C.

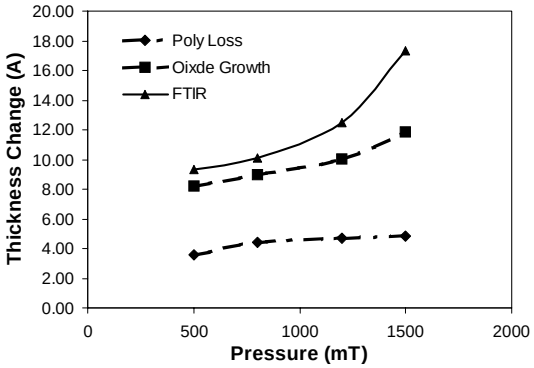


Fig. 6 Oxide growth and poly-Si loss as a function of pressure. Exposure time = 120s at 275°C.

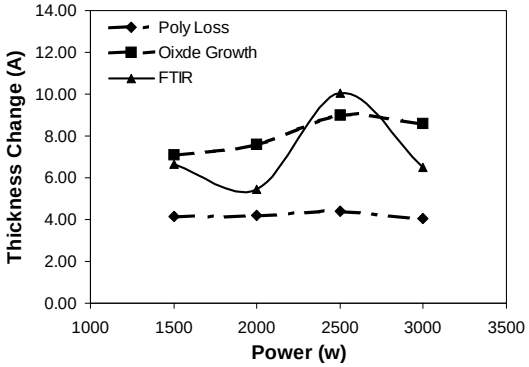


Fig. 7 Oxide growth and poly-Si loss as a function of RF power. Exposure time = 120s at 275°C.

O_2/FG Chemistry

Addition of FG to the O_2 chemistry is widely used in HDIS application due to its capability to remove residue on HDIS hardened crust over that of pure O_2. It has been shown that the high ash rate can still be maintained with up to 20% of FG content [7]. Results displayed in Fig. 9 show that the oxide growth from the O_2/FG mixture is larger than that of pure O_2 whereas poly-Si loss is independent of FG addition. In practice, O_2/FG process also has an additional advantage over pure O_2 process for process time reduction.

Ex. 1016, Page 694

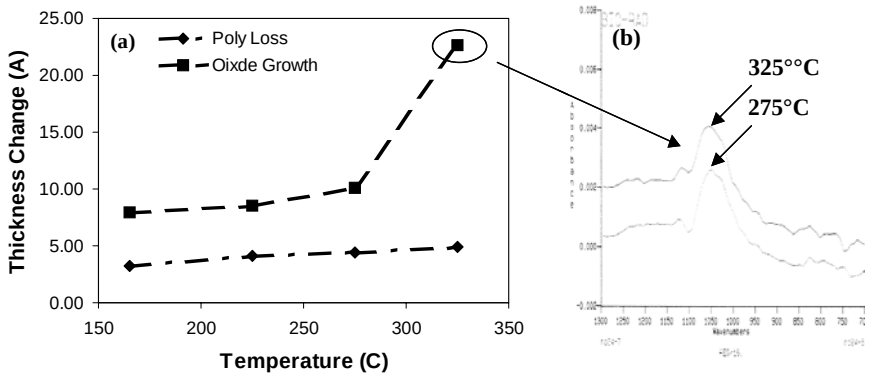


Fig. 8 (a) Oxide growth and poly-Si loss as a function of pedestal temperature and (b) FTIR spectra at 275 and 325 °C. O_2 plasma, exposure time=120s.

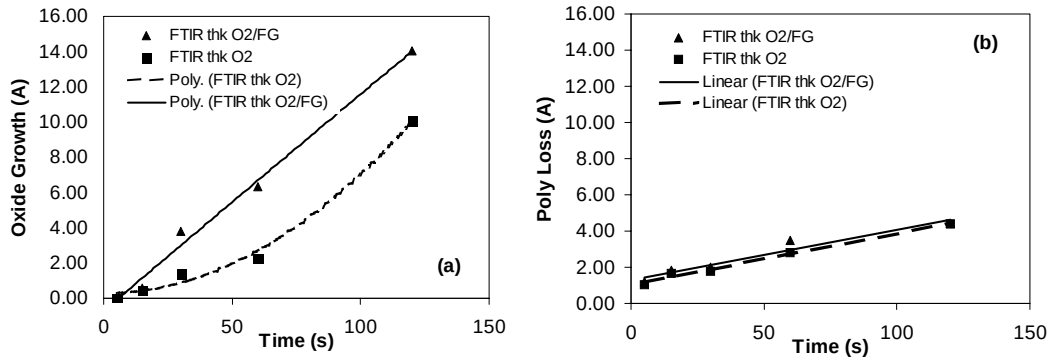


Fig. 9 (a) Oxide growth (normalized from FTIR intensity) and (b) poly-Si loss as a function of exposure time in 15% FG/O_2 plasma at 275 °C.

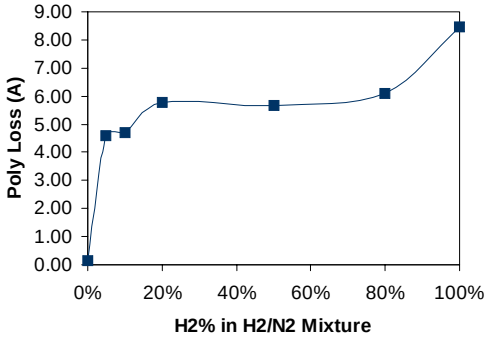


Fig. 10 Effect of H_2 percentage in N_2 plasma on poly-Si loss. Exposure time=120s at 275°C.

H_2 /N_2 Chemistry

H_2/N_2 chemistry outperforms O_2 chemistry on residue removal. Initial investigation of IR absorption showed no oxide or nitride formation (Fig. 4(b)) and no Si-H bonding after H_2/N_2 plasma exposure [8]. However, study of H_2 effect in H_2/N_2 gas mixture shows that poly-Si loss increases quickly with increasing H_2 content in H_2 /N_2 gas feed before saturated at ~20% H_2. The ploy-Si loss is independent of H_2 content between 20-80%,

which then has ~3Å loss from 80% to 100%, as summarized in Fig. 10. It is worth noting that there is a significant increase of poly-Si loss at around 6Å with 20% to 80% of H_2 in comparison to the 4.5Å loss from O_2 or O_2/FG chemistry. Since zero loss was observed in N_2 exposure, the loss of poly-Si is caused by H_2 addition.

Conclusion

Simultaneously minimizing oxide growth and poly-Si loss is a key challenge for FEOL process. Multiple passes of HDIS during gate formation introduce significant oxide growth and poly-Si loss. Oxide growth and poly-Si loss have been found to be dependent strongly on the gas chemistry and the exposure time. Traditionally used O_2 and O_2/FG chemistries yield the least poly-Si loss, up to ~ 4.5Å with pedestal temperature at 275 °C and 120s exposure time. Poly-Si loss has been found to be enhanced with the addition of H_2, from a maximum 6Å in H_2/N_2 to a maximum 13Å in H_2/O_2 within the stable H_2 percentage range, which is insensitive to the process pressure, RF power, and the pedestal temperature. Poly-Si loss also increases with the exposure time, independent of the chemistry used. It is also found that the ellipsometer used is acceptable for general trend studies, but not very reliable for evaluating oxide growth from a very short process time, high pedestal temperature, and high H_2 content. On the other hand, FTIR has been identified as a reliable methodology for charactering ultrathin oxide films. Oxide grows rapidly with extended exposure time and the addition of H_2. Slight increase of oxidation is observed with increasing process pressure, power, and pedestal temperature. Although a spectrum of chemistries are available for HDIS application, O_2 and O_2/FG chemistries still have their intrinsic advantage in terms of low poly-Si loss and low oxide growth even with extended process time. H_2/N_2 may be used when significant reduction of process time is required.

Acknowledgments

Authors would like to thank David Dutton, Neal Holmlund and Sing-Pin Tay for their support and helpful discussions.

References

1. B. Chan, B. Pemg, L. Sheu, Y. Chiu, and H. Tao, *IEEE 8th International Symposium*, V24-25, 73 (2003).
2. S. Rounds, A. Srivastava, and K. Han, *Semiconductor Manufacturing*, August, 28 (2006).
3. J. Borland, A. Mineji, W. Krull, M. Tanjyo, R. Hillard, and T. Walker, *Solid State Technology*, 49, 47 (2006).
4. J. Lambers and P. Hess, *J. Appl. Phys.*, **94**, 2937-3941 (2003).
5. X. Tan, J. Wojcik, and P. Mascher, *J. Vac. Sci. Technol.*, A 22(4), 1115 (2004).
6. A. del Prado, F. L. Martínez, I. Mártil, and G. González-Díaz, *J. Vac. Sci. Technol.*, A 17(4), 1263 (1999).
7. S. Xu and L. Diao, *J. Vac. Sci. Technol.*, A 26(3), 360 (2008).
8. J. Lambers and P. Hess, *Thin Solid Films*, **400**, 111-115 (2001).

ECS Transactions, 18 (1) 669-676 (2009)
10.1149/1.3096519 © The Electrochemical Society

Via Resistance Variation Improvement Through Amine Versus Fluorine Based Chemical and O_2 Dry Flushing for BiCMOS Technology

P.F. Tee**,** H.L. Teo, S.F. Liew, S.F. Chong, D.G. Lee, S.F. Lee

X-FAB Sarawak Sdn. Bhd., 1 Silicon Drive, Sama Jaya Free Industrial Zone, 93350 Kuching, Sarawak, MALAYSIA

The paper discussed the via resistance variation using different cleaning chemicals known as the amine and fluorine based chemical for 0.6um BiCMOS technology. Single and chain via resistance showed severe drift with fluorine based clean recipe especially at the wafer edge where the failure rate is approximately 28% for chain via resistance and 12% for single via resistance. The baseline chemical, amine based showed low failure rate of less than 3%. Resistance drift is mainly due to the high Ti/TiN polymer residues which are remained at the bottom of the via holes. This is because 0.6um technology tend to generate more by-products since the etch amount is higher compared to other advanced technology which have smaller critical dimension (CD) size. Focus of the paper is mainly on the recipe optimization with fluorine based chemical where the objective is to have comparable via resistance performance with the amine based chemical. Recipe was optimized for different wet cleaning tool, which includes Batch Spray Type and Single Wafer Type. All optimise recipes showed comparable via resistance performance with the amine based chemical. Apart from that, a study on via resistance by adding O_2 flushing etch step was conducted. The study was conducted on dual plasma source chamber and single plasma source chamber for BiCMOS technology. Adding O_2 flushing step helps to scavenge some of the polymer on the wafer. Single plasma source chamber with O_2 flushing step has better performance on production recipe compared to dual plasma source chamber.

Introduction

In integrated circuit technologies, BiCMOS, refers to the integration of bipolar junction transistors and CMOS technology into a single device. There have been two major approaches to BiCMOS process design. The low-end BiCMOS process incorporates a moderate-speed bipolar transistor into an aggressive CMOS process. The high-end BiCMOS process extends an aggressive bipolar process to include MOS device [1]. The difference between these technologies with other more advance technology is that 0.6 um BiCMOS technology has wider CD size. Therefore, etch amount for this technology is higher at via etch step and thus, more Ti/TiN polymer residue is generated at the bottom of the via holes. If cleaning recipes are not optimised, severe resistance drift can be observed due to Ti/TiN residue which are not entirely removed at the bottom of the via holes. Two significant resistance parameters were used for the comparison of the resistance performance which includes the chain via resistance and the single via resistance.

Ex. 1016, Page 697

In aluminium-based multilevel interconnect fabrication, via-holes are etched into the inter-metal dielectric layer for making electrical connection between the upper and lower metal layers. After via etch, it is critical to completely remove sidewall polymer residue generated during via etching to have a reliable metal filling process and good contact resistance [2-3]. Figure 1 shows an X-SEM picture of polymers trapped at the via-holes.

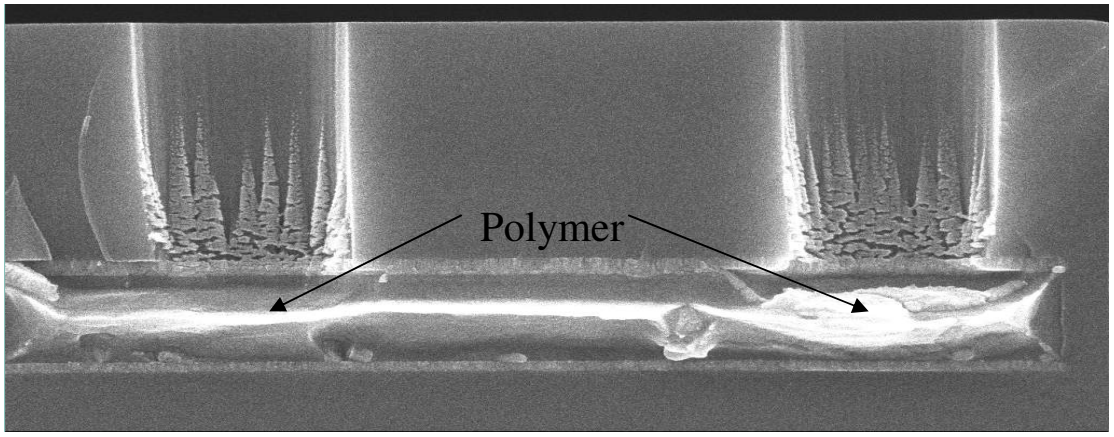
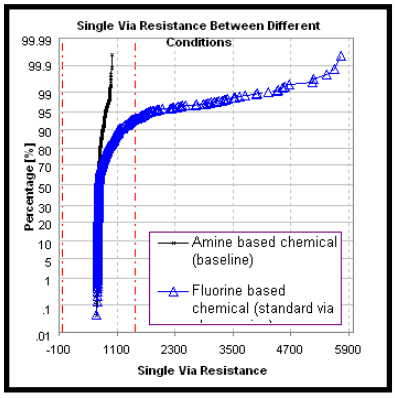
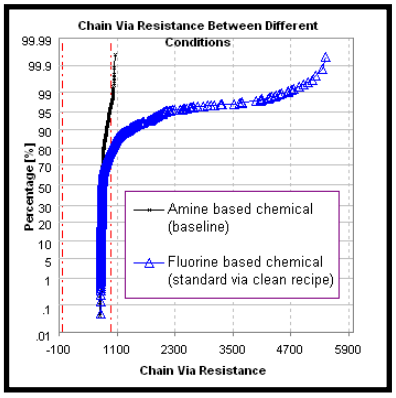


Figure 1: X-SEM picture of polymers trapped at the bottom of via-holes.

Amine based chemical is the baseline cleaning chemical for 0.6 um BiCMOS technology. It has better cleaning efficiency as compared to the standard via recipe; with fluorine based chemical. Figure 2 shows the different performance between amine and fluorine based chemical in single and chain via resistance.



(a) Single Via Resistance



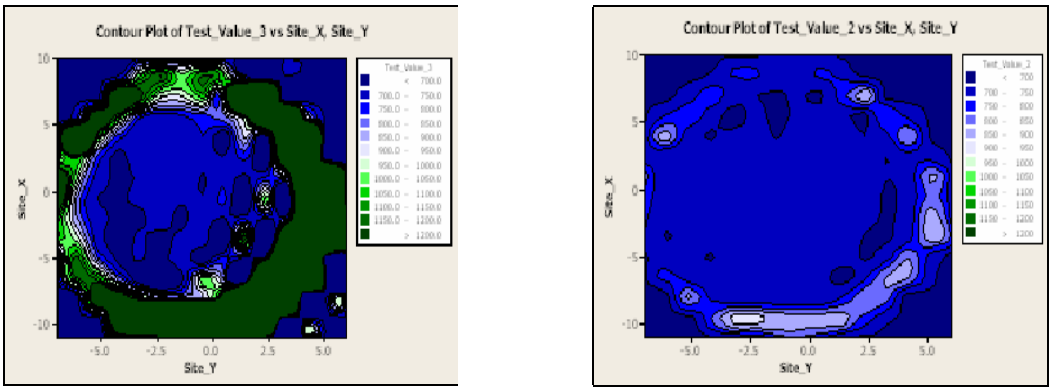
(b) Chain Via Resistance

Figure 2: Single and chain via resistance performance between amine and fluorine based chemical.

Single and chain via resistance showed severe drift with fluorine based clean recipe, where the average failure rate is 28% for chain via resistance and 12% for single via resistance. High failure rate can be observed especially at the wafer edge as shown at Figure 3. Figure 3 shows the contour plot of chain via resistance between fluorine and amine based chemical.

A study was also conducted on dual plasma source chamber and single plasma source chamber for BiCMOS technology. By adding O_2 flushing step, it will help to scavenge some of the polymer on the wafer.

Ex. 1016, Page 698



(a) Fluorine based chemical (b) Amine based chemical

Figure 3: Contour plot of chain via resistance for different cleaning chemical types.

Experiment & Simulation

Eight-inch wafers were used for all the experiments. The stack layer consists of AlCu/Ti/TiN/TEOS and 1.06 um resist as a patterning layer. The via was etched using a C_4F_8/Ar chemistry.

Simulations were done by optimizing recipe with fluorine based chemical where the objective was to have comparable via resistance performance with the amine based chemical. Recipe was optimized for different wet cleaning tool, which includes Batch Spray Type and Single Wafer Type. Split table for different cleaning chemicals and tool type are shown in Table 1.

Condition	Cleaning Tool Type	Split Condition	Recipe
A	Dipping Type	Dip for A minutes in amine based chemical	Baseline Recipe
B	Single Wafer	B times dispense (C minute soaking time for each dispense)	Standard Via Clean Recipe
C	Single Wafer	B times dispense (D minutes soaking time for each dispense)	Optimized Recipe
D	Spray Type	F cycles chemical dispense (E minutes dispense for each cycle)	Optimized Recipe
E	Spray Type	G cycle chemical dispense (E minutes dispense) plus soaking time of F minutes	Optimized Recipe

Table 1: Split table between amine based and fluorine based chemical at different tool types.

Basically, the simulations were done at 3 different wet cleaning tools which include the dipping type for amine based chemical, batch spray type as well as single wafer type for fluorine based chemical. Existing standard via clean recipe is not effective enough to remove the remaining Ti/TiN at the bottom of the via holes which will cause severe via resistance variation. Thus, simulations were carried out to optimize the cleaning recipe at single wafer type and batch spray type with fluorine based chemical.

Ex. 1016, Page 699

Currently, a study was also conducted on dual plasma source chamber and single plasma source chamber for BiCMOS technology. By adding O_2 flushing step, some polymer on the wafer will be scavenged away. Split table between dual plasma source chamber and single plasma source chamber is shown in Table 2.

Condition	Etch Tool Type	Cleaning Tool Type	Split Condition	Recipe
F	Single plasma source chamber	Dipping	without O_2 Flushing + Dip for A minutes at amine based chemical	Baseline Recipe
G	Single plasma source chamber	Dipping	O_2 Flushing + Dip for A minutes at amine based chemical	Optimized Recipe
H	Dual plasma source chamber	Dipping	O_2 Flushing + Dip for A minutes at amine based chemical	Optimized Recipe
I	Single plasma source chamber	Batch spray type	O_2 Flushing + C cycle of fluorine based chemical dispense (E minutes dispense) plus soaking time of F minutes	Optimized Recipe
J	Single plasma source chamber	Single spray type	O_2 Flushing + B times of fluorine based chemical dispense (D minutes soaking time for each dispense)	Optimized Recipe
K	Dual plasma source chamber	Dipping	without O_2 Flushing + Dip for A minutes at amine based chemical	Optimized Recipe

Table 2: Split table between different etch chamber and cleaning tool type.

After processing, wafers were sent for X-SEM examination. Electrical performances for all the different split conditions were analyzed. Test was conducted using single and chain via resistance comparison.

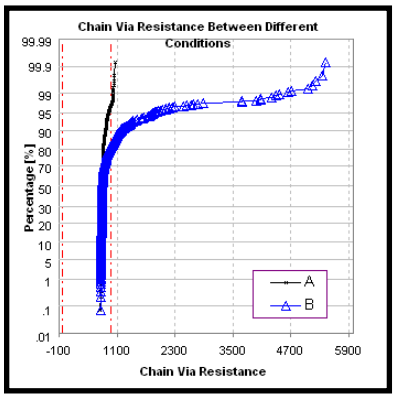
Results and Discussions

The simulation result is shown at the figures below. Figure 4 shows the baseline performance for amine based chemical in dipping type tool, as well as the standard via clean recipe at the single wafer type tool.

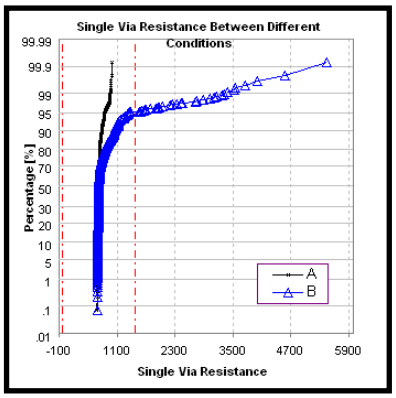
Via resistance performance for amine based chemical (condition A) shows a satisfactory result where all via resistance results are within the specification. Specification for chain via resistance is 1-1000 ohm whereas the specification for single via resistance is 1-1500 ohm. Standard via clean recipe (condition B) is not effective enough for 0.6 um BiCMOS technology due to the high generation of Ti/TiN polymer at the bottom of via holes. Polymer was formed due to the composition of the C_4F_8/Ti/TiN etch chemistry; sputtered metal material from the bottom of via is incorporated into sidewall deposited. Two times chemical dispense with 1 minute soaking time for each chemical dispense at single wafer type tool is insufficient enough to remove the entire polymer.

Thus, recipe was optimized (condition C) from condition B by increasing the soaking time from 1 minute to 4 minutes for each chemical dispenses. Results for condition C is shown in Figure 5. Condition C performance, is comparable with the baseline chemical.

Ex. 1016, Page 700



(a) Chain Via Resistance

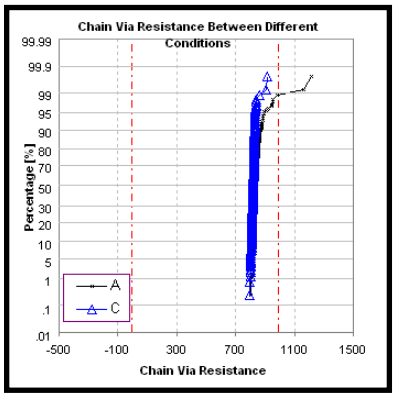


(b) Single Via Resistance

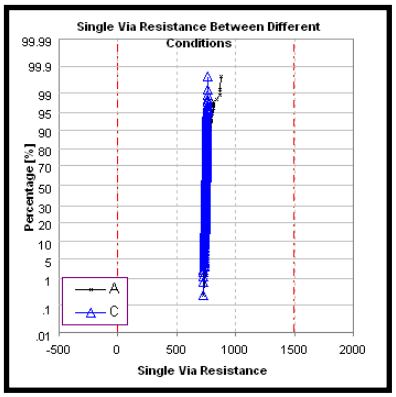
Figure 4: Performance for chain and single via resistance with the existing production recipe.

Longer soaking time is necessary in order to increase the reaction time between the polymers with the fluorine based chemical.

Apart from that, two cleaning recipes were also optimized using fluorine based chemical at the batch spray type tool which is condition D and condition E. Condition D has two times chemical dispense with a duration of 7 minutes for each dispense. Condition E has shorter process time because the process only involves one time chemical dispense for 7 minutes, plus soaking time of 5 minutes.



(a) Chain Via Resistance

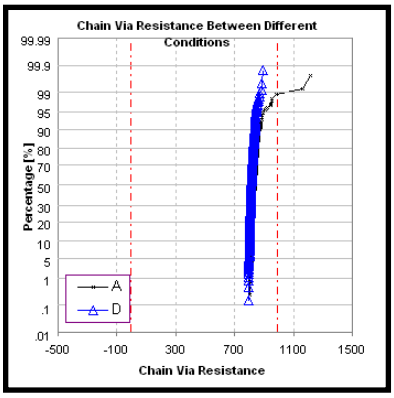


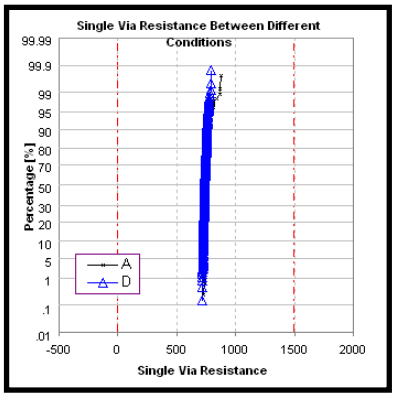
(b) Single Via Resistance

Figure 5: Performance for chain and single via resistance with optimized recipe at single wafer type tool.

Results for condition D and E is as shown in Figure 6 and Figure 7. Results for both condition D and E are comparable with the baseline via resistance performance. Comparison of via resistance performance between different conditions is shown in Figure 8. All conditions except condition B have comparable chain and single via resistance with the baseline performance.

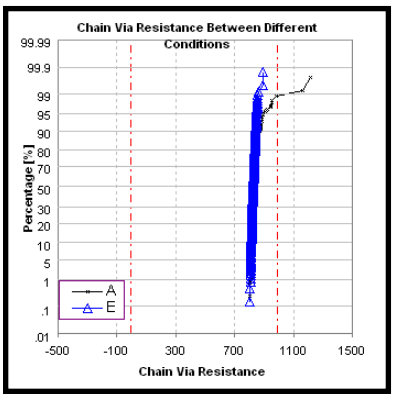
Adding O_2 flushing step in etch recipe helps to reduce the polymer generated. It helps to reduce the failure rate by 2% for chain via resistance especially at the wafer edge. Figure 9 shows the comparison in the via resistance by adding and without O_2 flushing step.

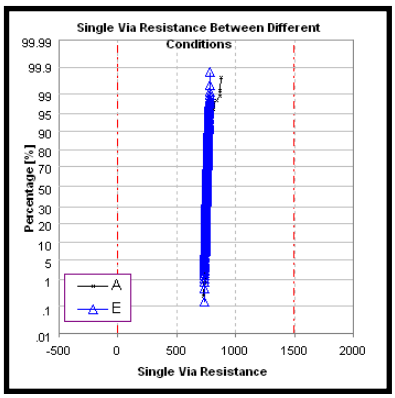




(a) Chain Via Resistance (b) Single Via Resistance

Figure 6: Performance for chain and single via resistance with optimized recipe at batch spray type tool with 2 times chemical dispense.





(a) Chain Via Resistance (b) Single Via Resistance

Figure 7: Performance for chain and single via resistance with optimized recipe at batch spray type tool with 1 time chemical dispense.

Single plasma source chamber with O_2 flushing step has better performance compared to dual plasma source chamber. Dual plasma source chamber with O_2 flushing step shows severe drift in chain and single via resistance. The result can be seen from Figure 10.

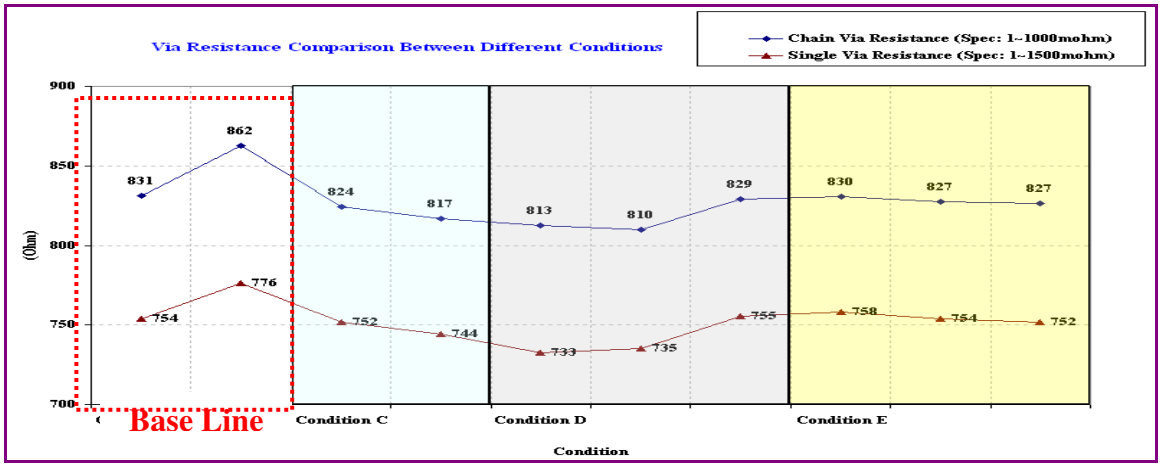
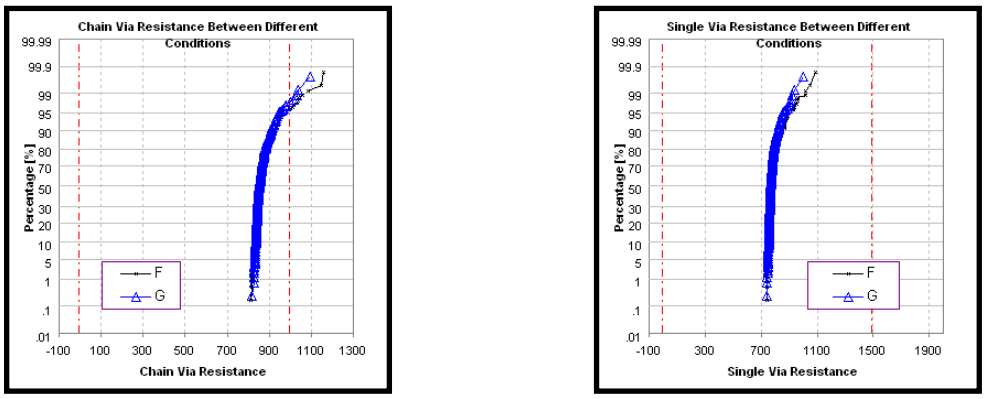


Figure 8: Comparison between different optimized conditions with fluorine based chemical with the baseline amine based chemical.

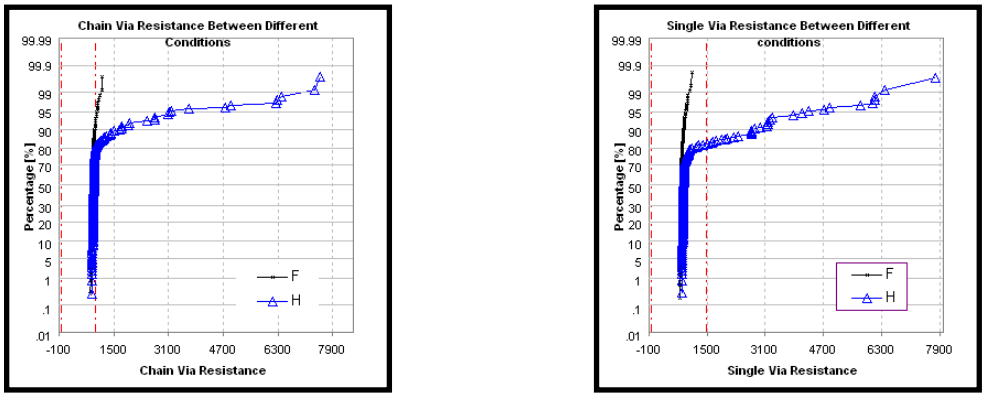
Apart from that, simulation was also conducted with dual plasma source chamber, with and without O_2 flushing step. As a result, both conditions showed severe drift for both single and chain via resistance.



(a) Chain Via Resistance (b) Single Via Resistance

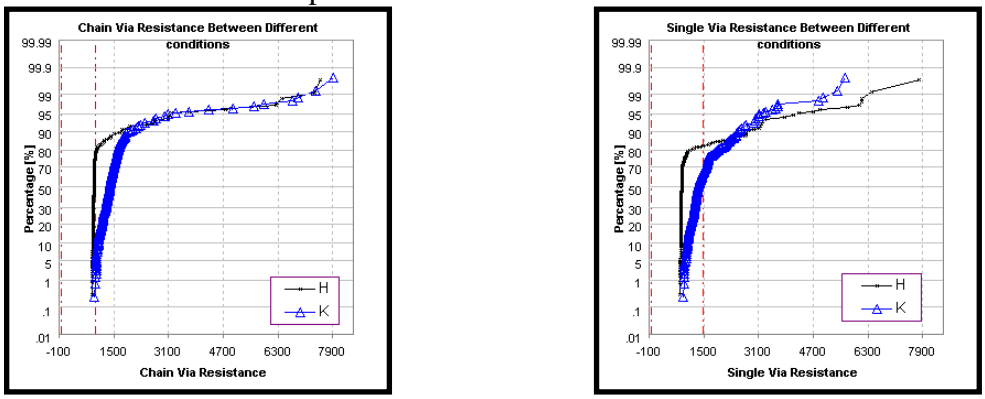
Figure 9: Performance for chain and single via resistance with and without O_2 flushing step.

This is due to production recipe at dual plasma source chamber was not optimized. Results for these conditions are as shown in Figure 11.



(a) Chain Via Resistance (b) Single Via Resistance

Figure 10: Via resistance comparison between different etch chamber.

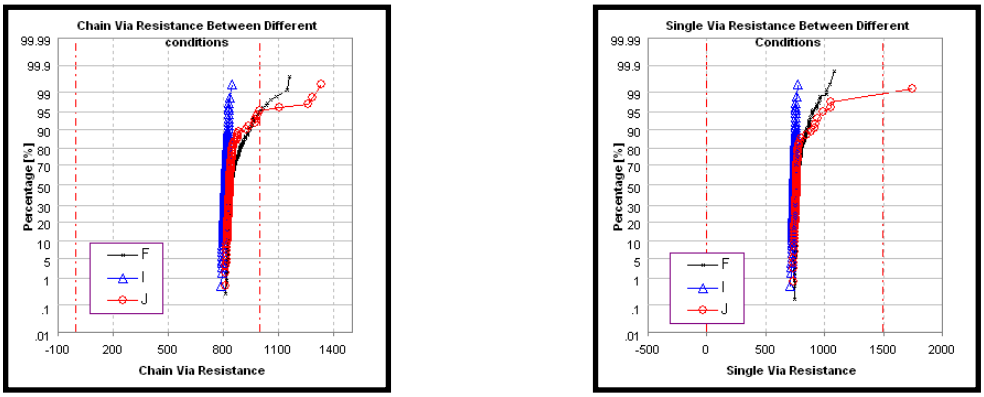


(a) Chain Via Resistance (b) Single Via Resistance

Figure 11: Via resistance comparison with and without O_2 flushing step at dual plasma source chamber.

Since single plasma source chamber has better performance than dual plasma source chamber, comparisons between the next few conditions (condition I and J) was done at single plasma source chamber with O_2 flushing but with different cleaning tool type (batch spray type and single wafer type).

Figure 12 shows the via resistance comparison between same etch chamber (Single plasma source chamber) but different cleaning tools. From the result showed, condition I has the best result compared to condition F and J.



(a) Chain Via Resistance (b) Single Via Resistance

Figure 12: Via resistance comparison between same etch chamber but different cleaning tools.

Conclusions

This paper is mainly focus on the recipe optimization by using fluorine based chemical in reducing the chain and single via resistance variation. High resistance variation is mainly due to polymers which are still remaining at the bottom of Ti/TiN. Recipe was optimized for 2 different types of cleaning tools which include the batch spray type and single wafer type. All the optimized recipes have comparable via resistance performance with the baseline amine based chemical. Besides, adding O_2 flushing step at via etch recipe at single plasma source chamber helps to scavenge some of the polymer on the wafer and thus, helps to reduce the failure rate by 2%.

Acknowledgments

We would express our gratitude to the X-Fab Management Team for sponsoring and supporting this paper. We would also like to extend our humble appreciation to Mr. J. L. Liu and Mr. K. A. Mohammad both from Etch for their technical advice and our dear friend Miss S. Png from Defect Management for her technical contributions to this paper.

References

[1] C.Y. Chang and S.M. Sze, “ULSI Technology”, Taiwan, 1988, pp. 92&502.
[2] Y. Wang, S. W. Graham, L. Chan and S. Loong, *J. Electrochem Soc.*, 144,1997, pp. 1522.
[3] P. J. Ireland, “Thin Solids Films I”, 1997, pp. 304.
[4] Boumerzoug, H. Xu, R. Bersin, L. Mikus, A. Horn, and D. Dopp, “Dry Process For Polymer Sidewall Residue Removal After Via-Hole Etching”, 2000, pp. 281-282.

Ex. 1016, Page 704

ECS Transactions, 18 (1) 677-680 (2009)
10.1149/1.3096520 © The Electrochemical Society

A Novel Approach for Electro Chemical Plating (ECP) Process Swirl Defect Reduction

Ri-Hui Sun[a], Ji-Wei Zhang[a], Jian-Yong Jiang[a], Yi-Jun Bian[a], Paul-Chang Lin[a], Cheng Xing[a], Xiao Li[b], Hua Zhou[a] *

[a] SMIC (Semiconductor Manufacturing International Corporation)

[b] Novellus System Inc.

* Correspondent author

Abstract

In copper metallization of BEOL, ECP is used for copper deposition in via or trench for interconnection after PVD barrier & seed process. But ECP film easily suffers pits defect with swirl pattern that is a killer defect to CP yield. This paper introduces a method to greatly reduce swirl defect by improving the wetting ability between Cu seed film and ECP chemicals.

Introduction

Copper has been considered to replace Al metallization in ULSI (ultra large scale integrated) circuits because of its lower resistivity and potentially higher resistance to electro migration. Typical approach for Cu deposition are copper barrier/seed layer deposition by advanced Physical Vapor Deposition (PVD) technique such as SIP, SIP EnCoRe etc, and subsequently bulk deposition by ECP (electro chemical plating). [1]

The PVD Cu seed film is very unstable and easily undergoes surface structural evolution even at room temperature when exposed to air. [1] The unfavorable surface evolution can cause poor wetting ability between Cu seed film and ECP chemicals during ECP process, which induces pits defect with swirl pattern to occur. For fresh PVD Cu seed film, more than 15% of production wafers suffered serious swirl defect that is a killer defect to CP yield.

The wetting ability can be improved by various surface treatments, such as pre-rinse. However, the above-mentioned surface treatment will cause other bad effects on production quality. In this paper, we introduce a new method to treat Cu surface, i.e. by air-dry spin of PVD Cu seed wafers prior to ECP plating process.

Ex. 1016, Page 705

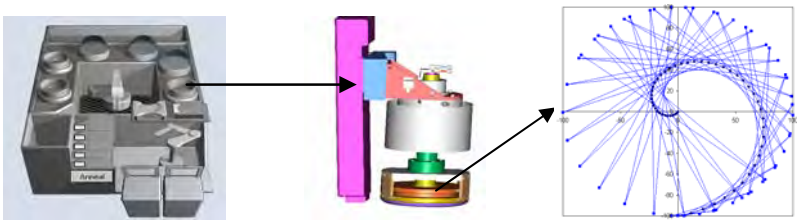
Experiment

Design of Experiment

Swirl pits are caused by the poor wetting ability in ECP process. To improve the wetting ability, wafer is dried spin in the air before plating process start, and then post CMP defect is checked.

Experiment Method

1) Novellus ECP SABRExT system in standard process condition Fig.1 (a).
2) AMAT CMP tool Mirra-mesa.
3) KLA-Tencor inspection tool 2351.



(a) (b)

Figure.1 ECP cell structure (a) and wetting/de-wetting track (b)

Wafer is deposited about 1000-2000Å Cu films as seed layer with about 150~350Å TaN/Ta as a barrier layer to prevent Cu diffuse into underlying dielectric. Novellus ECP tool SABRExT is used to perform plating process. KLA-Tencor inspection tool 2351 is used to scan wafer defects after CMP (chemical mechanical polishing) process. Figure.1 (b) shows that the swirl pattern is the wetting/de-wetting track.

Result and Discussion

Normally, PVD Cu film is deposited at much lower temperature than room temperature in order to avoid Cu grain agglomeration. [2] The PVD Cu film is very unstable and will undergo surface structural evolution/oxidation even at room temperature when exposed to air. But for fresh Cu seed wafer stored in POD, the oxidation will be restrained because of less oxygen.

ECP process is performed in $CuSO_4$ solution. So the wafer surface condition is very important to decide the quality of the process. Due to fresh wafer has poor wetting ability with $CuSO_4$ solution, when wafer exposes to air to do plating, the fresh seed wafer will hold bubbles in wafer surface. During ECP process, the bubbles not releasing in time will block the gap-fill and form holes in the Cu metal. Then, pits defect is found after CMP polishing process. During ECP process, wafer is rotated with certain entry angle to touch $CuSO_4$ solution, so the wetting/de-wetting track will form a swirl pattern Fig.1 (b). Thus, pits defect caused by bubbles along the wetting/de-wetting track will form swirl pits Fig.2.

Ex. 1016, Page 706

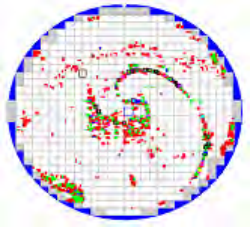


Figure.2 typical swirl defect map

Generally, pre-rinse the wafers before plating is a widely used effective method to eliminate swirl pits. But for pattern wafers, the rinsed DI water will enter via and trench and cannot be removed completely due to limited rinse dry which will cause seed layer to be damaged and lead to some unexpected side effect to CP yield. Fig.3 shows pre-rinsed wafers have about 2% CP yield loss compared to air-dried spin wafers.

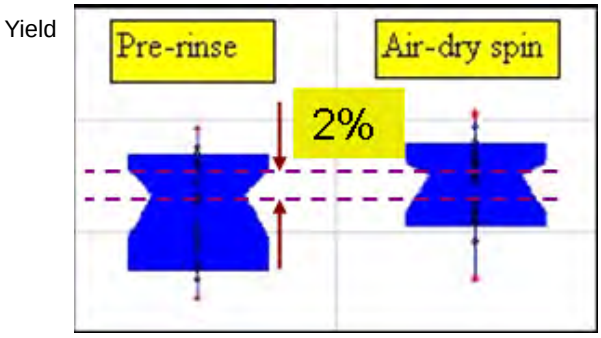


Figure.3 Yield data of pre-rinse VS air-dry spin

Compared to fresh seed wafer, the seed wafer stayed in air for a long time which surface is oxided with oxygen has better wetting ability due to surface energy is released for CuOx is easier to dissolve into acid solution than pure copper. Base on such concept, a method of air-dry spin, which can increase the speed and chance for copper oxidation, is used to improve the seed wafer wetting ability with $CuSO_4$ solution. In process, wafer is put into the plating cell of SABRExT tool and rotates by cell motor above the chemicals for certain time, then wafer downs into $CuSO_4$ solution to complete the following plating process with normal condition. With the air-dry spin function and the improved wetting ability between Cu film and ECP chemicals, the defect count is greatly reduced and swirl pits is eliminated.Fig.4

Ex. 1016, Page 707

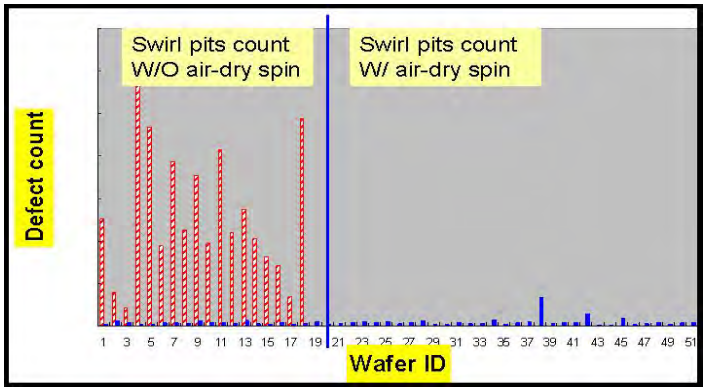


Figure.4 Comparison of defect count before/after the implement ion of dry-spin

Conclusion

Pre-rinse is proved of the effective methods and widely used to reduce defect in ECP plating process. But it will bring some unfavorable side effect to CP yield while air-dry spin has no impact to CP yield. Air-dry spin is performed in SABRExT tool plating cell and no other additional instrument is needed. The spin time and rotation speed can also be flexibly adjusted to get high throughput. So dry spin some time with certain rotation speed in plating cell before ECP plating will greatly reduce swirl defect without other side effect.

Acknowledgment

Authors would like to thank our supervisors Allan_Zhou, Division head of SMIC-S1-TF and Daniel_Fu, Center head of SMIC-S1 center, for their encouragement and guidance.

References

1. Jeng, C.C. et, al., BEOL process integration of 65nm Cu/Low k interconnects. Interconnect Technology Conference, 2004, Proceeding of the IEEE 2004 International.

2. Peijun Ding, et, al.,. Advanced Cu barrier/seed development for 65nm technology and beyond. Solid-State and Integrated Circuits Technology, 2004. Proceedings. 7th International Conference, 2004.

CHAPTER 7

PACKAGING TECHNOLOGY

ECS Transactions, 18 (1) 683-688 (2009)
10.1149/1.3096521 © The Electrochemical Society

Material Technologies for 3D Semiconductor Packaging

N. Takano, T. Tanaka, S. Katogi, A. Nagai

New Applied Materials R&D Center, Hitachi Chemical Co., Ltd.,
48 Wadai, Tsukuba-shi, Ibaraki 300-4247, JAPAN

The demand of three dimensional PKG (3D PKG) effective for decreasing the mounting area with higher performance is rapidly increasing around mobile electronic devices. The form of the PKG is evolving variously depending on the usage. In this paper, we introduce the materials, material design technologies and the combination of them such as Die Bonding Film for Stacked Multi-Chip PKG, Epoxy Mold Compound and Substrate for Package on Package (PoP) and Underfill Material for Advanced Flip-Chip Package.

Introduction

Mobile devices (cellular phone etc.) are becoming miniaturizing with higher performance. The semiconductor packages (PKGs) mounted on the limited area of substrate in these devices are also becoming smaller and thinner day by day. Therefore, the mount technology had evolved from current SMT (Surface mount technology) to three dimensional mount technology [1]. Moreover, the advancement of System in Package (SiP) technology which mounts two or more chips in one PKG is also remarkable to satisfy the demand of improving the processing speed and power consumption [2]. To meet the requirement, not only each characteristic but also the combinations of materials for PKG are becoming very important. In this report, we introduce new material technologies and the combination suitable for the three dimensional mounting and SiP technologies.

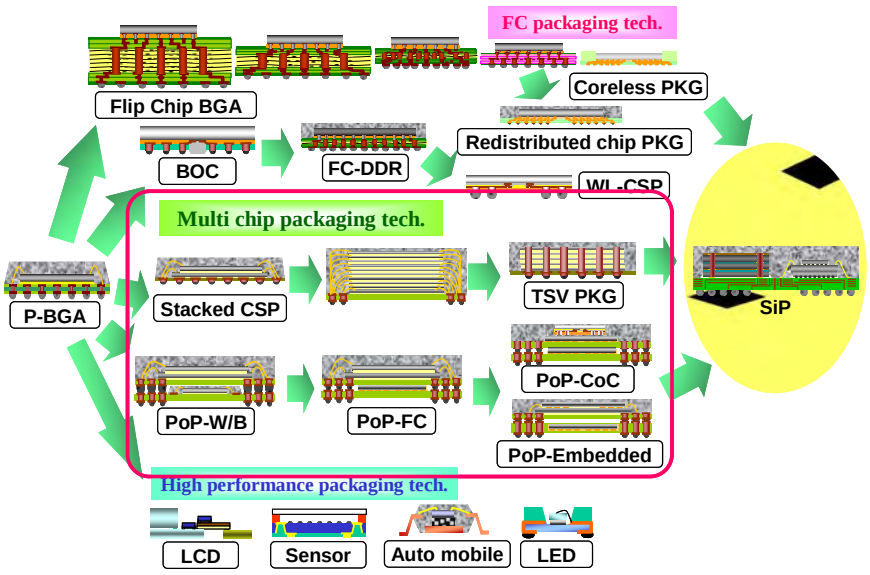


Figure 1. Trend of Semiconductor Packaging

Ex. 1016, Page 711

Packaging material line up of Advanced Packages

Advanced packages (Including 3D PKG) have various materials (Figure 2) such front end materials as Low K, Buffer coat and CMP slurry, and back end materials as Die Bonding Film, Paste, Epoxy Mold Compound, Liquid encapsulant, Substrate, Solder Resist and etc. (Figure 2). With these materials, we can fabricate a variety of the advanced PKGs.

We evaluate the compatibility of these materials for customers. At the same time, for each complex PKG, we analyze the stress simulation to obtain the improvement index of each material. If there are some problems, we give the evaluation data and index to our R&D department. The R&D department supplies improvement materials with evaluation data. So, customers can accelerate their development. Especially, in the case of 3D PKG, the excellent matching of properties of materials is very important. We will propose the combination of the best materials to customer.

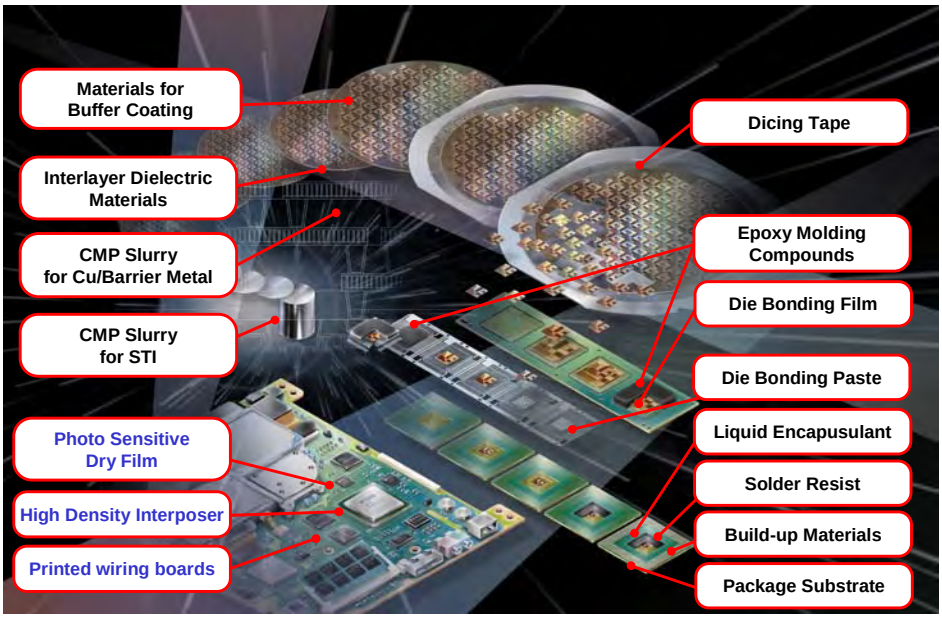


Figure 2. Products Line-up of Advanced Semiconductor PKGs

Advanced Material Technologies for 3D PKG

Die bonding Film for Stacked Multi-Chip PKG

Stacked multi-chip PKG widely noticed for memory devices is under further development in the field of semiconductor packages getting smaller and thinner with higher performance.

Typical structure and tendency of stacked multi-chip package are shown in Figure 3. In this case, the property of Die Bonding Film (DAF) is very important to enhance the PKG performance. Day by day, the thickness of the wafer is decreasing, and the number of stacked chip is increasing. Accordingly, the following problems have been picked up (Figure 3b).

/ The warpage of thin wafer after laminating DAF
/ The warpage of PKG after attaching chip to the substrate
/ Delamination and die crack during thermal cycle test

Ex. 1016, Page 712

To solve the problem, we have developed novel polymer alloy system used for DAF. This material has achieved a low elastic modulus and excellent heat resistance, and is effective to decrease the warp and the thermal stress of stacked CSP [3].

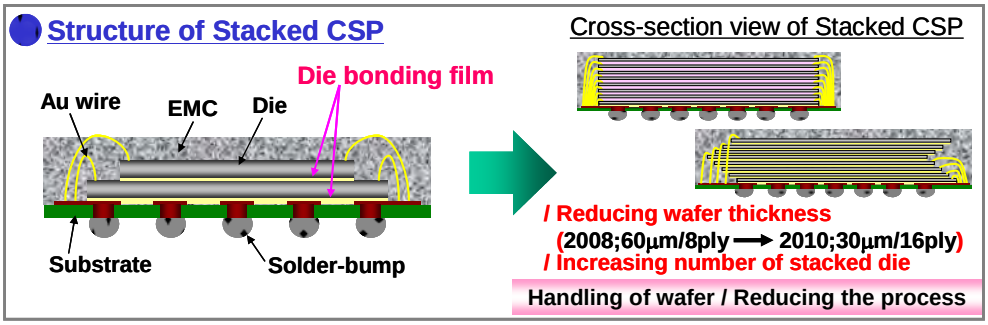


Figure 3a. Typical structure and Main Issues of Stacked Multi-chip PKG

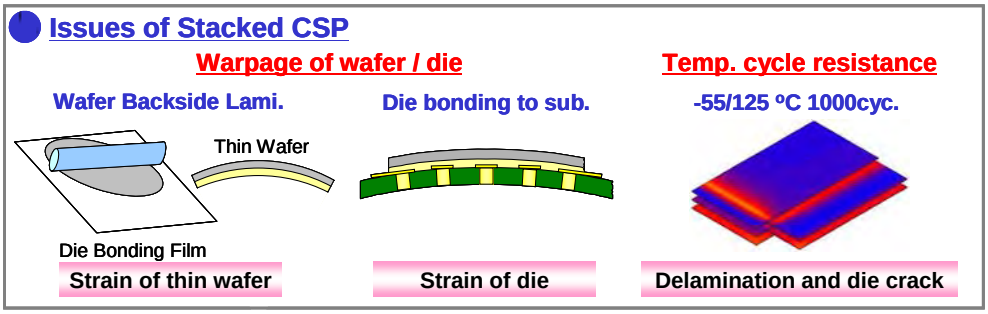


Figure 3b. Typical structure and Main Issues of Stacked Multi-chip PKG

Epoxy Mold Compound and Substrate for Package on Package (PoP)

PoP is a mounting form that accumulates PKG of one or more chips. In general, PoP stacks the memory package on the logic package to save the space on the PCB. Because the height of total package increases in PoP, it is necessary to thin the thickness of the substrate and the molding material as much as possible (Figure 4). The bothering problem of thinner PKG is the connection trouble of PoP especially when the warpage behaviors of the top and the bottom PKG are different. So, it is important to control or decrease the warpage of each PKG [4]. As the thermal expansion properties of the Substrate, Epoxy Mold Compound (EMC), Underfill material (UF) influence the warpage of PKG directly, the adjustment of the properties becomes more important than before (Figure 5).

Table 1. Trend of Package on Package

Items	2006	2008	2010
Bump pitch / mm	0.65	0.50	0.40
PKG size / mm²	14×14	12×10	10×10
Mold thickness / mm	0.30	0.20	0.13
Bump height / mm	0.33	0.23	0.18
Wire diameter / mm	18	15	15

Ex. 1016, Page 713

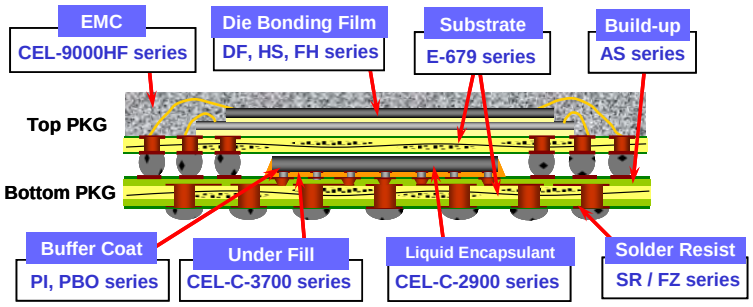


Figure 4. Typical Structure of Package on Package and Related Materials

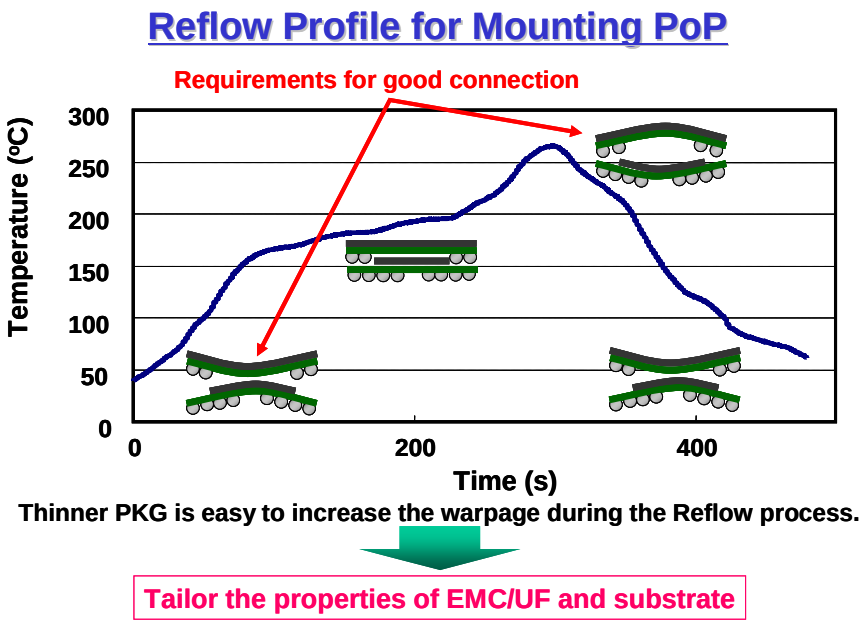


Figure 5. Warpage issue of PoP in reflow process

The coefficient of thermal expansion (CTE) of the substrate and EMC influence the warpage of PKG directly. We have developed a low CTE substrate and a high CTE EMC suitable for thin PKG by using new resin system [5].

Underfill Material for Advanced Flip-Chip Package

As shown in table 2, the trend of flip-chip package is going towards higher density, narrower gap between chip and substrate, larger chip size, and higher speed (lower K). As for UF, the low stress and narrow gap filling will become more and more important in the future. In addition, the pre-applied material such as non-flow underfill (NUF) and UF film will be demanded strongly for ultra-narrow gap PKG and through silicon via (TSV) PKG. We are developing the novel highly tough resin system for the demand. These materials are useful for reducing the crack of large and Low K chip. The newly developed UF film will be especially suitable for reducing some steps of complex processes.

Ex. 1016, Page 714

Table 2. Trend of FC-PKG and UF properties

Silicone chip trend		2006	2008	2010	2012	Challenge of UF
Chip Parameter	Bump count	6,000	10,000	14,000		➢Narrow gap filling ➢No voiding ➢NUF, UF film --- New technology
	Bump pitch	170um	150um	120um		
	Die gap	75um	50um	35um		
	Die size	15mm□	20mm□	30mm□		➢Low warpage ➢Low stress to chip
	Die thickness	600um	400um	200um		
Low-K		K=3.5	3.0	2.5		
Pb free bump		Pb-Sn		Pb free		Reflow performance
Solder joint		Flux cleaning		Non clean flux		Flux compatibility

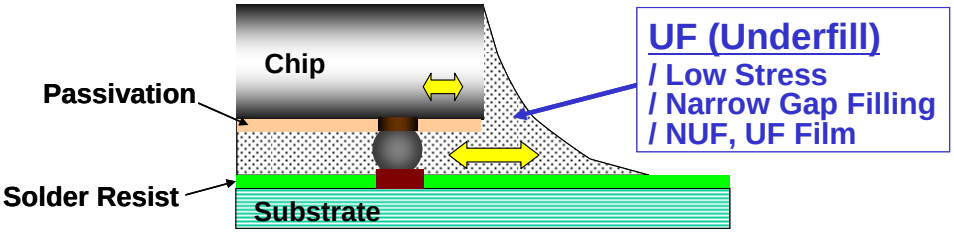


Figure 6. Under Fill materials in the future

Advanced Materials for the Future Packaging

New Substrate for 3D Mounting; We have developed new substrates for ultra-thin multilayer boards. Main features of the material are 1) excellent static bending ability, 2) high heat resistance, 3) excellent dimensional stability for multilayer process [6]. By using these materials, you can materialize thin and bendable multilayer boards for three dimensional mounting PWB (Figure 7). We have two kinds of substrate (Cute series ; with the glass fabric and 5000I series ; ultra thin film).

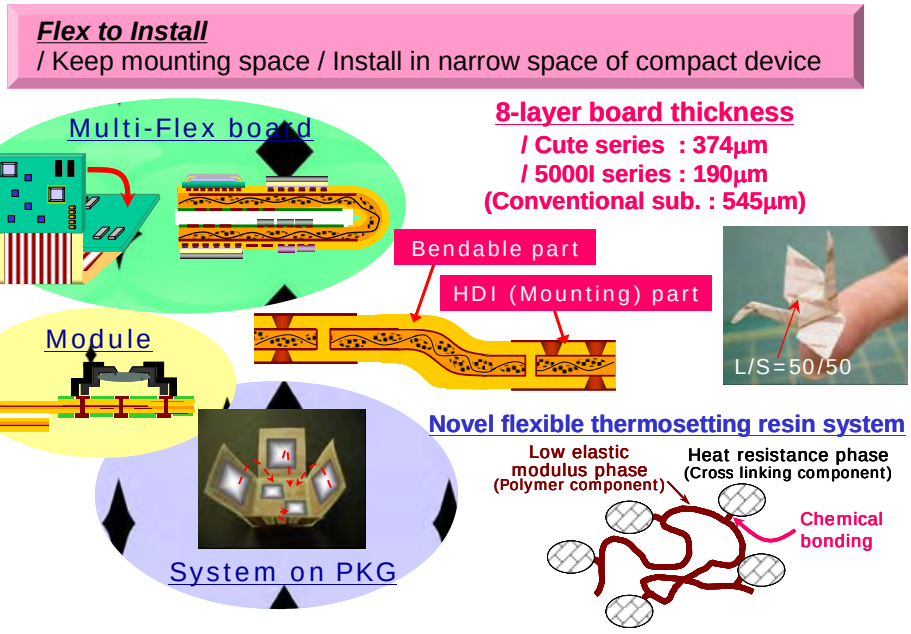


Figure 7. New Substrate for 3D Mounting

New Conductive Paste; We have developed a new conductive paste suitable for low temperature reflow process [7]. This novel paste is composed of metal of low melting temperature and high heat resistant thermosetting resin system. It forms the metallic bond and the resin fillet that reinforces the interconnection after 150°C reflow process (Figure 8). The metallic bond decreases the connection resistance and the resin fillet improves the connection reliability in TCT. By using the paste, you can easily decrease the reflow temperature for reducing the warpage of ultra-thin PKG.

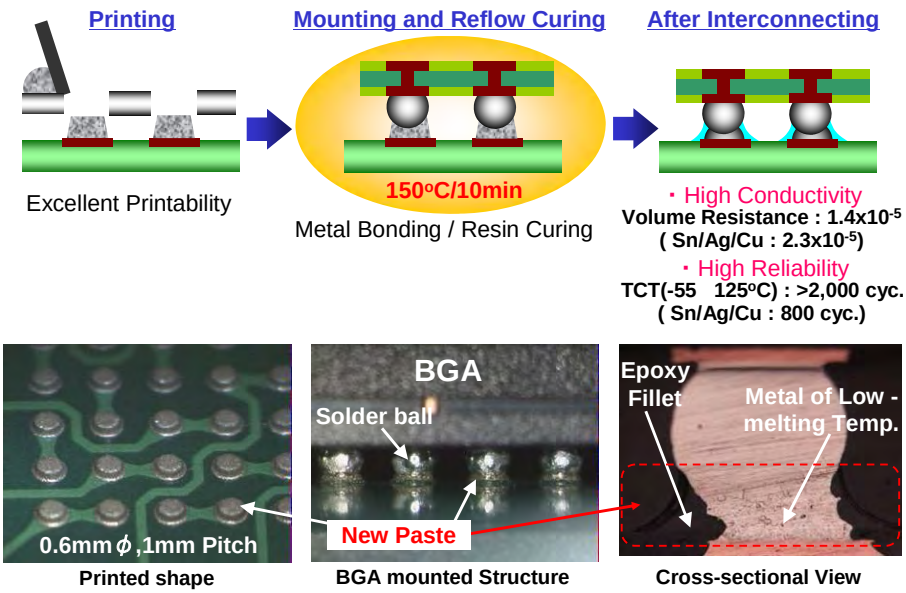


Figure 8. New Conductive Paste for Low Temperature Reflow Process

Conclusions

SiP, 3D PKG and the integrated PKG with MEMS and optical technology will be widely used in various fields of electronics.

In this paper, we introduced the line up of products of Hitachi Chemical for the wide variety of the advanced PKG and outlined the novel materials and technologies for PKGs such as Die Bonding Film for stacked multi-chip PKG, Substrate and Epoxy Mold Compound for PoP, Underfill material for advanced flip-chip package, new substrate for 3D mounting and new conductive paste for low temperature reflow process.

References

1. T. Tanaka, *Advanced Electronic Packaging Technology Symposium*, p. 1 (2008)
2. S. Ouchi, *Electronics Jisso Gakkai-shi*, vol. 10, No. 5, p. 368 (2007)
3. T. Matsuzaki, *Hitachi Chemical Technical Report*, No.46, p. 39 (2006)
4. Y. Takagi, *Denshizairyo*, vol.1, p. 66 (2008)
5. K. Masaoka, *Denshizairyo*, vol.1, p. 70 (2008)
6. N. Takano, *Hitachi Chemical Technical Report*, No.47, p. 21 (2006)
7. H. Hayashi, *Micro-Electronics Symposium, MES 2008*, p. 175 (2008)

Ex. 1016, Page 716

ECS Transactions, 18 (1) 689-694 (2009)
10.1149/1.3096522 © The Electrochemical Society

High-Speed Cu-TSV for 5:1 *i*TSV Applications

R. Beica, P. Siblerud

Semitool Inc., Kalispell, Montana 59901, USA

Transition from two-dimensional (2D) to three-dimensional (3D) integration seems to be critical for achieving miniaturization and increased performance and an unavoidable trend for the semiconductor industry. Because of the large variety of applications and requirements as well as a equipment and processes available today for 3D Interconnect, there are still questions about choosing the most suitable integration solutions that can not only provide the performance and advancements required for next generation electronic devices, however a cost effective implementation as well.

This paper will address 3D integration using copper electrodeposited through silicon vias (TSV) technology. Various process parameters, such as via profile, seed layer uniformity, wettability and chemical stability as well as process parameters and equipment design were found to be critical in successfully filling TSV structures using copper electrodeposition. This paper will describe the effect of these factors with a focus on the challenges and costs associated with interconnect type TSV structures (*i*TSV).

Introduction

The traditional way for IC integration using two dimensional approaches has been proven to have several limitations when it comes to building more complex electronic devices, with increased functionality and performance while reducing the form factor and cost. Industry experts estimate that traditional scaling of devices will reach its physical limits within the next 10-15 years [1]. Although, significant efforts continue to be spent in further integration of copper and low-k dielectrics, the industry has also started in developing new and more advanced packaging technologies.

One of the most advanced technologies today, which is continuously gaining a lot of interest in the semiconductor industry, is 3D integration using TSV technology. By providing shorter connections with TSV, the distance of information flow on a chip will be reduced (by up to 1000x). TSV connections will also allow addition of increased number of channels or pathways (up to 100x more than 2D chips), necessary for information flow. Less efficient wire bonds can also be replaced by shorter and more robust TSV electrodes. The main advantages of building vertical packages using 3D integration include miniaturization of size and weight, ability to integrate heterogeneous technologies in the same package, shorter connections therefore faster speed, reduction of parasitics and power consumption [2-5].

Ex. 1016, Page 717

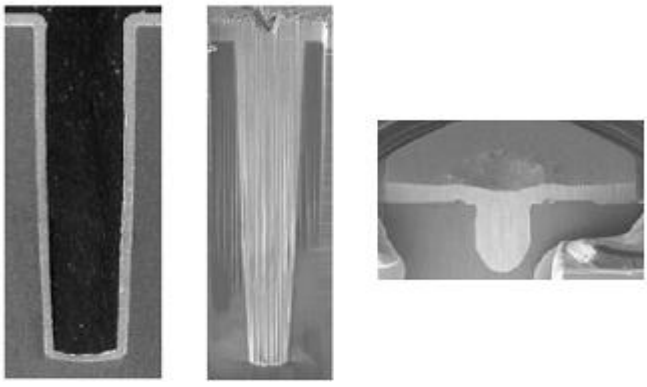
A significant amount of work has been performed in the past few years to better understand the benefits and challenges of 3D integration for various packages and applications and develop processes and integrated solutions that will enable smoother and faster adoption of this technology. One of the main concerns still facing adoption is the integration cost of TSV.

Copper Electrodeposition for TSV

Different materials and processes have been proposed and tested for via filling. The selection of the most suitable materials and methods of forming these TSV connections is usually done based on performance requirements, process considerations, type of application, reliability and cost of integration.

The most common material used for via filling and formation of TSV electrode is copper. Copper can be deposited either though dry or wet methods. Dry methods, especially plasma vapor deposition (PVD) is usually applied and is currently the preferred choice of depositing the conductive seed layer on the sidewalls of the TSV features. A very common wet copper metallization is the electrodeposition process. Copper electrodeposition, a very well known method in the electronics and semiconductor industry, due to its history of successful applications for printed wiring boards (PWB), copper interconnect and various wafer level packaging (WLP) applications, and being a more cost effective way of metallization, it is considered the most suitable approach of forming the copper TSV eletrodes.

Depending on the application, and especially the feature sizes, there are typically three types of TSV copper electrodeposited processes, as shown in Figure 1:



(a) (b) (c)

Figure 1. Types of TSV deposition
(a) lining; (b) full-filling, and (c) full-filling with stud formation

If via lining is more practical for larger features, more commonly found in MEMS and sensor applications, full filling is applied over a wider range of via dimensions, ranging from interconnect features (*i*TSV type) to larger, packaging type vias (*p*TSV). To form additional connections, either as a redistribution line (RDL) or solder bonding (copper stud and lead-free solder), a photoresist patterns can also be added [6].

Ex. 1016, Page 718

Although copper electrodeposition is very well known, when it comes to TSV, there are certain challenges and process considerations that are critical for successfully implementation of this process for 3D integration. Via profile, seed layer uniformity, ability to wet high aspect ratio (AR) features, chemical and process stability, optimization of chemistry with process parameters are some of the most critical challenges associated with electrodeposition of copper TSV structures.

Seed Layer Uniformity

As with any integration, finding the most suitable unit processes and the best solutions to seamlessly integrate them is very important. When it comes to feature with more challenging dimensions and aspect ratios higher than 5:1 (such as the *i*TSV features), traditional PVD systems cannot provide the uniform and conformal seed layer coverage inside the via necessary for achieving super-conformal deposition. Discontinuous seed layer cannot provide the conductivity film required and uniformity for continuous nucleation and growth of copper inside the vias, resulting in poorly filled vias, as shown in Figure 2.

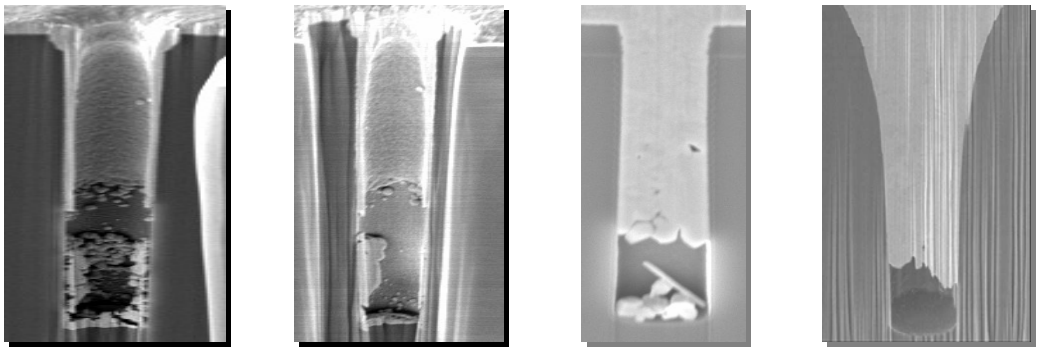
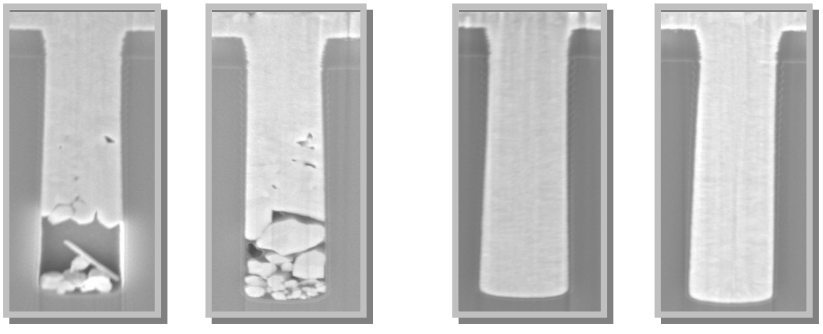


Figure 2. Non-uniform seed layer deposition on the sidewall and bottom of TSV features

Ionized PVD systems, although have the capability to go beyond 5:1 aspect ratios, are significantly more expensive. To overcome these limitations, wet chemical processes that can deposit or repair the seed layer have been developed. Successful application of such processes (as shown in Figure) have been obtained for feature with aspect ratios as high as 33:1.



(a) (b)

Figure 3. Seed layer repair on 2x11µm features: (a) bottom voids due to poor seed layer coverage; (b) void free plating after wet seed repairing process.

Easy control and application and successful implementation make wet seed approach a strong candidate for 3D integration, especially for more challenging and aggressive features.

Cost Effective Integration of TSV

Successful metallization of a wide range of TSV features for various applications can be performed today using copper electrodeposition, as shown in figure 4:

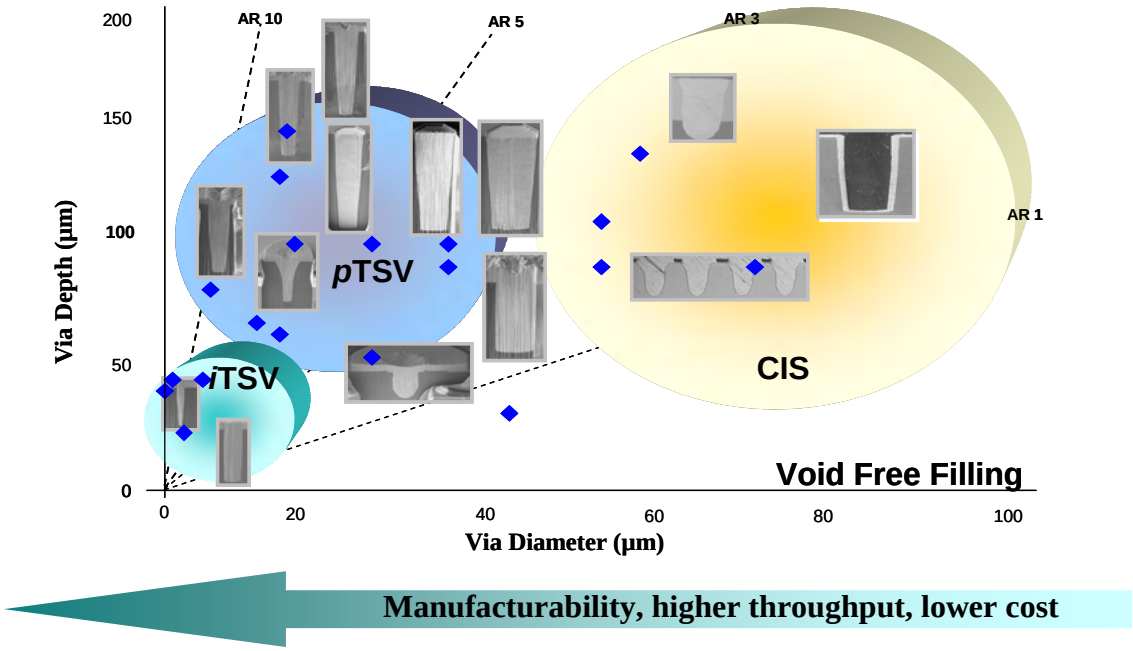


Figure 4. Wide range of void free TSV electrodeposition, from interconnect type feature (iTSV), to packaging type features (pTSV) and very large features used for CMOS image sensors [7]. Manufacturability, due to increased throughput and ultimately reduced cost of ownership favors smaller features.

Industry estimation of costs associated with 3D Integration using electrodeposited TSV technology shows that electrodeposition process can become quite costly when the volume of the vias are larger, such as packaging type TSV features. This is a result of the longer plating times and especially the cost of consumables associated with this type of features. Decreasing the feature sizes, currently considered by several TSV adopters, can significantly reduce the overall costs associated with this step, making smaller features more feasible from manufacturability point of view.

Equipment reliability and consumables are considered even more critical cost factors. With traditional plating equipment, material and chemical costs associated with plating can become significantly higher than capital costs, especially for larger type features and longer processes. In addition to cost, not having a stable system of controlling and maintaining the concentration of the each individual component of the copper baths could become detrimental to the final performance of the plating process, again, more significantly in the case of packaging type features, with longer processing times. State of the art equipment with high speed automation and membrane systems have been

Ex. 1016, Page 720

developed to address these challenges and to significantly reduce the consumable costs over traditional type, non-membrane systems.

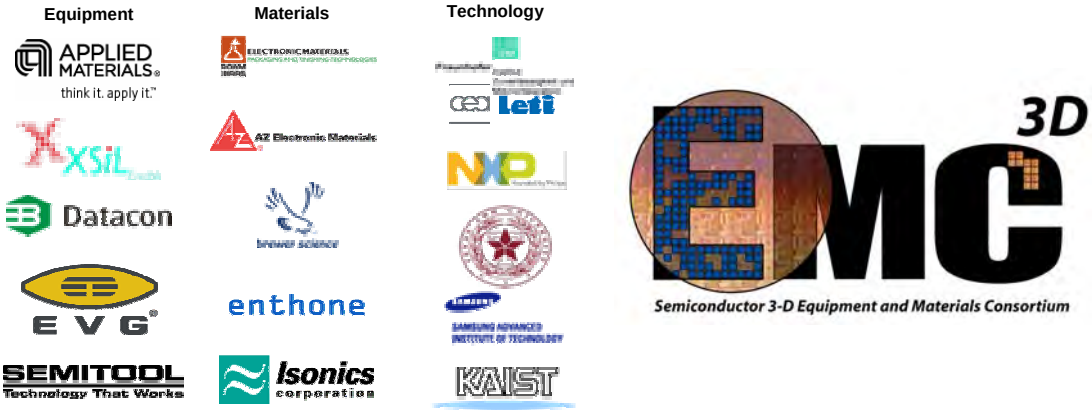
If form factor and performance are currently the market drivers for TSV development, cost has been identified to be the ultimum factor that will decide how broadly this technology will be adopted across the industry.

In order to lower the costs and, at the same time, solve the integration challenges associated with 3D integration using TSV, strong collaboration between supply chain members is critical. EMC3D consortium has been successful in developing advanced processes and integrated solutions and to continuously decrease the cost associated with copper TSV packages. Recent cost of ownership calculations, using current processing and capital costs, show that sub $190 per wafer can be achieved today with integrated 5x30µm via-first chip stacking technology and further developments are in progress to reduce even more the cost.

Acknowledgments

The authors would like to thank EMC3D consortium members for their continuous efforts to improve and achieve more effective TSV integration. EMC3D consortium includes the following member companies: Applied Materials (California, USA)

- XSiL (Dublin, Ireland) for laser drill and dicing
- Semitool (Montana, USA.) for wet processes (Cu TSV plating, Wet Etch/strip)
- EV Group (Schärding, Austria) for wafer/die alignment and stacking
- Fraunhofer IZM (Institute for Reliability and Mikrointegration, Chemnitz, Germany)
- LETI (Laboratoire d'électronique et de technologie de I'information, Grenoble, France)
- SAIT (Samsung Advanced Institute of Technology, Seoul, Korea)
- KAIST (Korea Advanced Institute of Science and Technology, Daejeon, Korea)
- TAMU (Texas A&M University, College Station, Texas)
- Rohm and Haas Electronic Materials PFT (Freeport, N.Y.)
- Enthone Inc. (West Haven, Conn.)
- AZ Electronic Materials (Somerville, N.J.)
- Brewer Science (Rolla, Mo.)



The authors would also like to thank to the members of the TSV group from Semitool, including Charles Sharbono, Marvin Bernt, James Burnham, Paul Kusler and Ross Kulzer for their continuous help with development and characterization of the TSV electrodeposition processes.

References

1. P. Garrou, *3-D ICs Enter Commercialization,* Semiconductor International, November Edition, 2008.
2. P. Siblerud, *Cost Effective Copper iTSV Integration,* Pan Pacific 2009
3. R. Beica, C. Sharbono, T. Ritzdorf, *Copper Electrodeposition for 3D Integration*
4. Solid State Technology, *IBM tips TSV 3D Chip Stacking Technique*, April 2007
5. D. Schmauch, J.M. Thevenoud, R. Beica, M. Wimplinger, *3D Packaging – How to Build 3D Packages from Design through Materials & Equipment*, Advanced Packaging, February 2008
6. D. Schmauch, B. Kim, and T. Ritzdorf, *3D Packaging Enabled with Electrochemical Deposition Techniques from Varied Electronic Industry Segments,* Pan Pacific Microelectronics Symposium, January 2006
7. R. Beica, P. Siblerud, *Advanced Electrodeposition Technologies for 3D Integration,* Pan Pacific 2009

ECS Transactions, 18 (1) 695-700 (2009)
10.1149/1.3096523 © The Electrochemical Society

Advanced Technologies for In-Line and Post-Processed TSV Integration.

Eric Beyne, Bart Swinnen, Jan Van Olmen, Deniz S. Tezcan

IMEC, Kapeldreef 75, 3001 Leuven, Belgium

Through Silicon Vias (TSVs) are an enabler for 3D system integration. A wide diversity of technologies is being proposed. Different application domains have different TSV requirements and justify different integration approaches. Key technologies are discussed for the in-line or post process realization of TSV connections. The impact of wafer thinning and 3D processing on 130nm CMOS devices is discussed.

3D technology and interconnect wiring hierarchy

3D system integration studies the possibilities to interconnect active devices in different 2D-planes. These interconnect can be considered at different levels of the wiring hierarchy, from the package, over the global to the local interconnect levels. As 3D moves down in interconnect hierarchy, the density of 3D interconnects will need to increase exponentially. The choice of the 3D interconnect level(s) has a significant impact on the system design and the required 3D technologies, resulting in a strong interaction between system design and technology requirements.

Classification of 3D Technologies

In literature one can find a very large variety of approaches to 3D die stacking. In an attempt to structure the discussions on 3D technologies, we propose to categorize these technologies based on the interconnect hierarchy levels they address. Interconnect levels were well defined by ITRS for on-chip [5] and Jisso for package level [6] interconnect levels. A particular advantage of this approach is that this corresponds to specific technology platforms matching the existing manufacturing infrastructure ("fab-type"): wafer-fab, wafer-level-packaging, traditional packaging and board assembly. We identify the following "flavors" of 3D-technology:[1-4]

- 3D-SIP: 3D interconnects realized at 2nd and 3rd Jisso level of the packaging interconnect hierarchy (package level), relates to the traditional packaging interconnect technologies such as wire-bonded die stacks and package-on-package integration
- 3D-WLP: 3D interconnects realized at 1st Jisso packaging level (Wafer or bondpad level). This technology relates to the wafer-level packaging infrastructure, such as RDL and bumping. A key characteristic of 3D-WLP is that the 3D interconnects are processed after the IC fabrication, 'post IC-passivation' (via last process).
- 3D-SIC (3D-Stacked-IC): 3D interconnects at the global or intermediate level of the chip wiring hierarchy (0th Jisso Level). Through Si-vias are realized in the wafer fab. Stacking and assembly can be done in the packaging companies.
- 3D-IC: 3D interconnects at the local level of the chip wiring hierarchy. In this case the stacking is also performed in the wafer fab.

Ex. 1016, Page 723

An advantage of this approach is that the roadmaps for the required 3D-interconnect pitch can be derived from the more general roadmap of interconnect density at any level of the interconnect hierarchy.

For 3D-WLP stacks, interconnects are realized at the wire-bond pad level of the devices, and therefore require a 3D interconnect density similar to the IC-bond pad pitch. Today this implies 3D-interconnect pitches down to 40µm, with a roadmap of pitch reduction down to about 20µm.

For applications that repartition the circuit blocks of a System-on-chip, SOC, in the 3rd dimension (3D-SOC [1]) 3D interconnect are to be realized at the level of the global interconnect chip layers. This 3D-SIC stacking technology requires significantly smaller TSV-pitches, scaling down from 10 µm to 1µm.

Going further to the lowest local interconnect layers, the actual 3D-IC technology, extremely tight 3D-interconnects are required, approaching the dimension of the transistor or gate-pitch. This requires a very different technical approach, basically stacking layers of transistors. Because of the limitations of wafer bonding with respect to overlay accuracy, such structures must be sequentially bonded and structured.

In this paper 3D-WLP and 3D-SIC TSV technologies are discussed in more detail.

3D-WLP, post-processes TSV processing

Post-processed TSV technologies typically require lengthy Si-etch and metal fill process steps. In order to obtain a cost effective TSV process, the TSV volume must be minimized. This can be done by scaling down the via diameter or thinning the Si wafer. Practically this requires a higher aspect ratio via. For aspect ratio's (Si thickness/diameter via) below 2-3 standard PVD techniques can be used to deposit the barrier metals and seed layers. Higher AR's, up to 5 and even 10 can still be done with enhanced sputtering tools at a higher cost of ownership. Very high AR's require CVD processes for metallic seed deposition.

Therefore in order to have a cost effective process for 3D-WLP TSV structures, a low AR has to be maintained, typically 2-3. For large diameter TSV structures in thick Si wafers, plating times for fully filled TSV's may approach 3 hours per wafer, resulting in an uneconomical process. In that case, it is preferable to use a conformal Cu-plating in an open TSV structure as shown in figure 1.[7] In this case the Cu-plating time can be reduced to a few minutes.

This conformal Cu-TSV is however not scalable to very tight interconnect pitches. For the higher density 3D-WLP TSV structures, a fully filled TSV structure is required. As the plating time is a critical factor in the overall TSV process cost, a thin Si-wafer is favored. Current thin die handling capabilities for post-TSV packaging allow for thinning the die to about 50µm.

The IMEC approach to a 3D-WLP fully filled Cu-TSV is shown schematically in figure 2. First the wafers are bonded to a temporary carrier wafer, followed by wafer thinning down to 50 µm. The next step is to etch a circular trench, "donut", in the Si, stopping at the front side oxide layer. This circular trench is about 5 µm in width. This circular trench is filled with a polymer dielectric using a spin-coating technique. The polymer layer is opened on top of the Si-center of the (polymer) donut. This allows for the selective (wet or dry) etching of the central Si pillar and the back-etching of the oxide at the bottom of the hole, exposing the metal contact pad. The final steps are the PVD deposition of a barrier and Cu seed layer, followed by the electroplating of the Cu via. An

important advantage over the sloped via process is that no lithography at the bottom of the via is required.

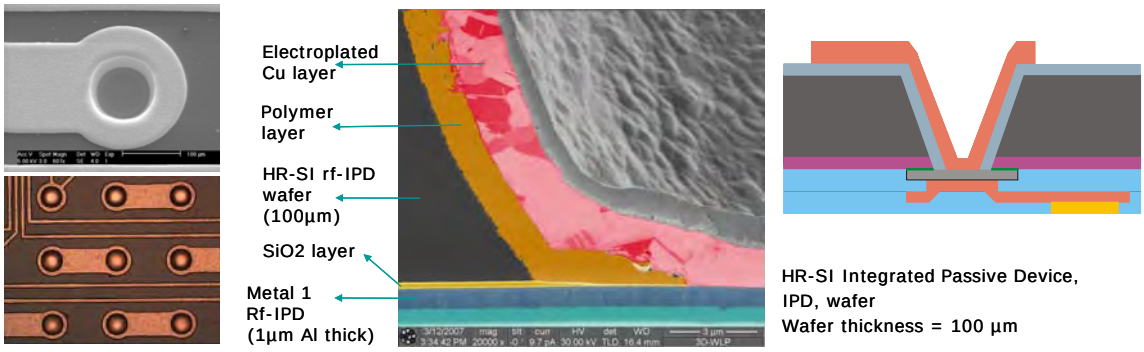


Figure 1: Through Si-via technology for relatively large via diameters and larger Si-wafer thickness, using a conformal polymer and metal coating of the low aspect ratio 1:1 via.

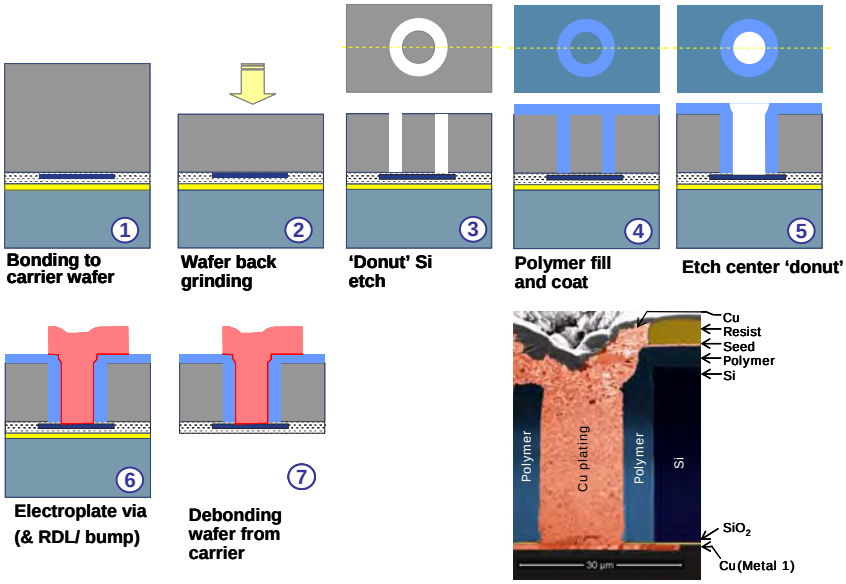


Figure 2: Through Si-via technology for small via diameters and small Si thickness, using an annular ring polymer isolation layer and a fully filled Cu via.

3D-SIC, in-line TSV processing

For 3D-SIC smaller diameter and pitch TSV structures are used. Our approach to 3D-SIC stacking, consists of realizing the Cu TSV, so-called "Cu-nail", during the IC process.[8-11] This Cu-nail is processed after the FEOL process (transistors), but before the BEOL process (multilayer damascene interconnect layers), as shown in figure 3.

The Cu-nail is realized by plasma etching a ±25µm deep Si hole with a diameter of 5µm. A modified Cu single damascene process is used to fill the hole. A CVD oxide layer is used as thin dielectric insulating layer and a CMP stop layer. A metal barrier and Cu seed is then deposited by PVD. Subsequently, the via hole is then filled with electroplated copper. CMP is used to remove the Cu "overburden". After this process, standard BEOL is used to finalize the Si-die.

For these TSV dimensions, the Cu-electroplating time can be significantly reduced, using state-of-the-art plating tools that allows for parallel processing of wafers, a high throughput can be realized. The RIE plasma etching process is however slower for these higher aspect ratio structures, making it a cost-critical process with respect to throughput.

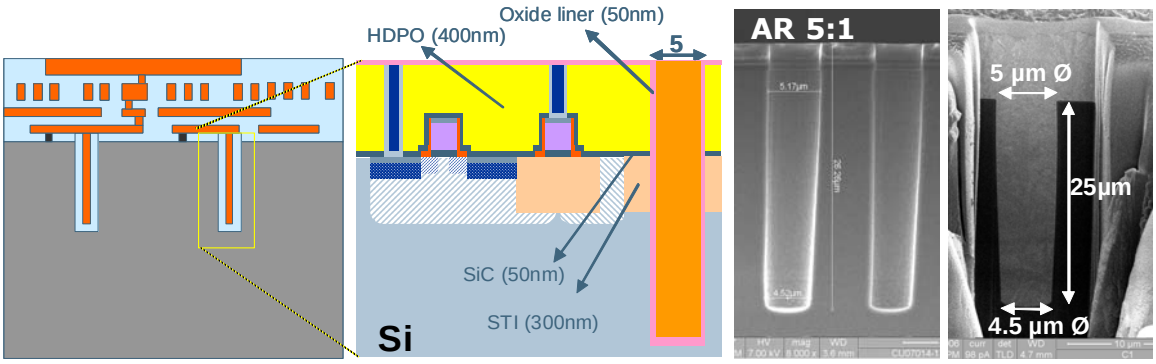


Figure 3: 3D-SIC Cu-nail process for realizing high density 3D-SIC TSV via connections.

Impact TSV technology on the Electrical Performance of CMOS devices

The impact of the reduction of substrate thickness, TSV integration and the stacking process on the electrical performance of both n- and p-type MOSFETs has been investigated in detail.[9] Parameters like saturation threshold voltages, intrinsic transistor performances (figure 4) of both n- and p-type MOSFETs have been monitored at various steps in the 3D-SIC flow.

Particular attention is given to the contacts to the active area and to polysilicon gate to insure the integrity CMOS FEOL after the TSV integration that follows the CMOS contact module. Gate leakage was found to be unaffected. The impact of grinding on the substrate regions has been evaluated previously [12]. Diode leakage currents, which can be taken as indicators for substrate damage and stress levels resulting from the thinning and stacking process by Cu–Cu thermocompression, do not show significant deviation from the reference performance. No major disturbance has been observed on the electrical characteristics of n- and p- type devices for a TSV-device distance larger than 1.5µm (figures 5-6), consistent with the simulations of [13].

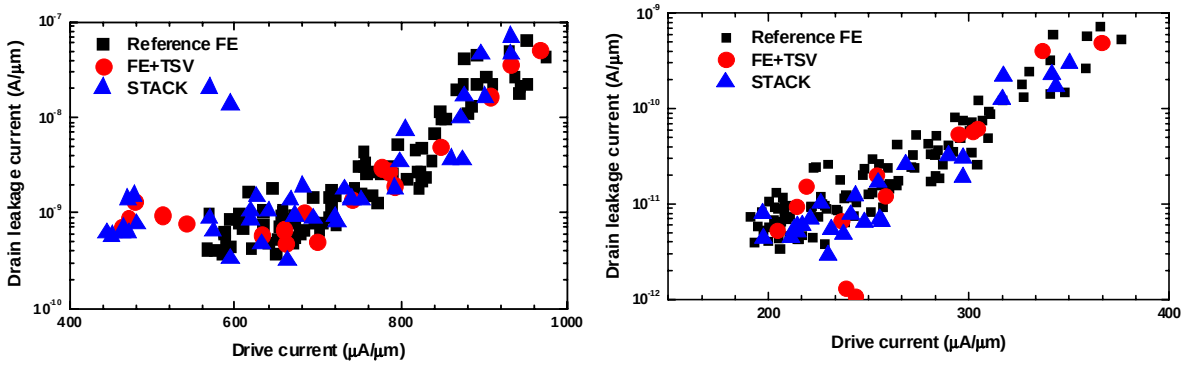


Figure 4: Comparison of the intrinsic transistor performance for both n- (left) and p-MOSFETs (right) on reference wafer (reference FE), a thinned front end wafer with TSVs (FE+TSV) and on top tiers of stacked dies (STACK).

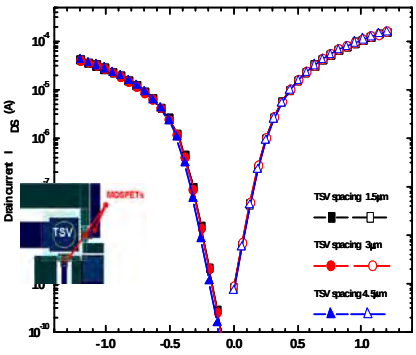


Figure 5: Drain current versus the gate bias in saturation for both n- and p- type MOSFETs at different distances from a TSV.

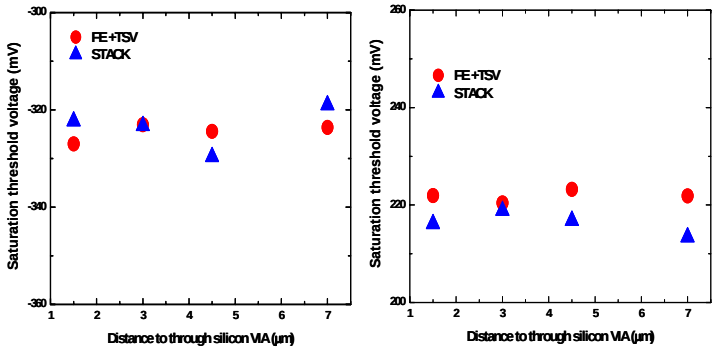


Figure 6: Comparison of the median saturation threshold voltage for both p- and n-type MOSFETs on a thinned front end wafer with TSVs (FE+TSV) and on top tiers of stacked dies (STACK) at different distances from a TSV.

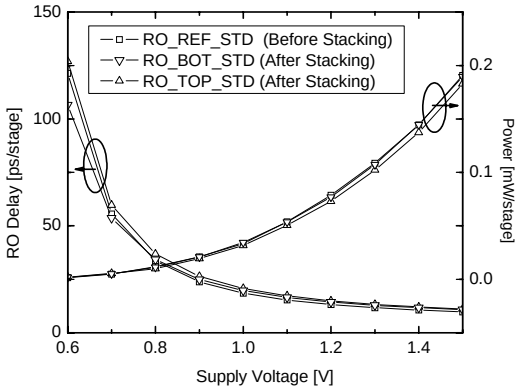


Figure 7: Delay & Power comparison for Top & Bottom Standard ROs and RO reference.

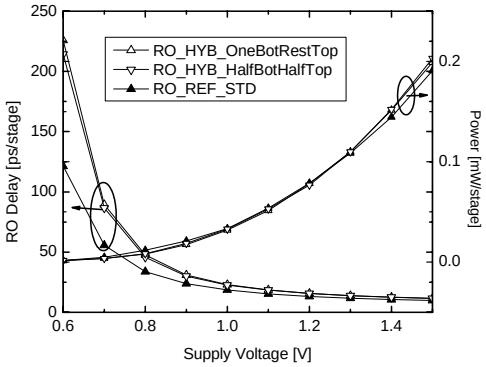


Figure 15: Delay & Power comparison for Hybrid and reference ROs.

To gauge the performance of FEOL on top and bottom strata along with the TSVs, 41 stages Ring Oscillator (RO) with varying configurations are designed and fabricated. The FEOL on both top and bottom strata are 7.5µm away from the edge of the TSV (Ø= 5µm). All ROs are followed by 8 frequency dividers (Divide by 256) and a buffer. RO without any load lying entirely on the bottom strata before stacking (RO_REF_STD) is taken as reference RO. Unloaded RO lying entirely on the top strata (RO_TOP_STD) and entirely on the bottom strata (RO_BOT_STD) measured after stacking have been compared with RO_REF_STD. Hybrid RO with 21 inverters on bottom strata and 20 on the top strata (RO_HYB_HalfBotHalfTop) connected through 2 TSVs at the end of top and bottom inverter chains have been compared with RO with one inverter at the bottom strata and rest of the inverter stages on top strata (RO_HYB_OneBotRestTop). Both designs consist of 2 TSVs interconnecting the inverters on top and bottom strata. Bottom ROs are designed such that they are measurable before as well as after stacking.

RO Delay variation and power per stage for RO_TOP_STD and RO_BOT_STD after stacking are compared with the reference RO_REF_STD in figure 7. Both top and bottom ROs show comparable performances when compared to the reference RO. Figure 8 compares the delay variation & power per stage of the hybrid designs with the reference RO_REF_STD. As expected, the 2 TSVs cause larger delay in the ring oscillator operations. Moreover, irrespective of their positions in the design, delay and the power are comparable in both hybrid designs.

Conclusions

3D interconnect technology follows the interconnect wiring hierarchy. Different technologies will address different types of 3D-integration at different levels of the interconnect hierarchy. In order to become successful they should be aligned with the manufacturing infrastructure.

For coarse-pitch, large size, 3D-WLP post processed. TSV structures, a conformal Cu coating is to be recommended. For smaller 3D-WLP vias, 20 to 30 µm diameter, fully filled Cu vias can be effectively realized, provided the wafer thickness is reduced to about 50 µm.

3D stacked Integrated Circuits with copper TSV First approach and D2D stacking process have been demonstrated. Integrity of top and bottom CMOS FEOL throughout the 3D-SIC flow has been proven. Impact of TSV proximity on active devices has been analyzed. Ring oscillator performance of the top, bottom and hybrid configurations before and after stacking confirm the communication between top and bottom dies interconnected through TSVs.

Acknowledgments

The authors acknowledge the IMEC 3D research team A. Mercha, G. Katti1, C. Huyghebaert, J. Van Aelst, E. Seppala, Zhao Chao, S. Armini, J. Vaes, R. Teixeira Cotrin, M. Van Cauwenberghe, P. Verdonck. K. Verhemeldonck, A. Jourdain, W. Ruythooren, M. de Potter de ten Broeck, A. Opdebeeck, T. Chiarella, B. Parvais, I. Debusschere, T.Y. Hoffmann, W. Dehaene, M. Stucchi, M. Rakowski, Ph. Soussan, S. Biesemans, R. Cartuyvels.

References

1. E.Beyne, *IEEE ISSCC*, 15-19 February 2004, San Francisco, pp.138-145 (2004).
2. E.Beyne, *IEEE IEDM*, 2-5 December 2001, Washington, D.C., S23-p3 (2001).
3. E.Beyne, *IEEE IITC*, 5-7 June2006, San Francisco, CA, USA, 1-5 (2006).
4. E.Beyne, *Semi Technology Symposium*, 7 December 2007, Makuhari, Japan, S9P1 (2007).
5. *The International Technology Roadmap for Semiconductors*, http://www.itrs.net/
6. *Jisso*: http://jisso.ipc.org/levels_p2.htm
7. D. Sabuncuoglu Tezcan et al, Proc. 57th *IEEE ECTC*, 29 May - 1 June 2007, Reno, NV, USA (2007).
8. B. Swinnen, et al., *IEEE IEDM*, 11-13 December 2006, San Francisco, (2006).
9. J. Van Olmen et al., *IEEE IEDM*, 15-17 December 2008, San Francisco, CA, USA, pp. 603-606 (2006).
10. E. Beyne, *4th Int. Symp. Microelectronic. Packaging*, 28 September 2005, Seoul, South Korea (2005).
11. A. Jourdain et al., *IEEE IITC*, June 2007, San Francisco, CA, USA (2007).
12. K. De Munck et al., *IEEE Electron Device Letters*, **29**, 4, (2008).
13. C. Okoro et al, *IEEE IITC*, June 2008, San Francisco, CA, USA, pp. 16-18 (2008).

ECS Transactions, 18 (1) 701-706 (2009)
10.1149/1.3096524 © The Electrochemical Society

3D DRAM in Server Application

Biao Cai, Vipinchandra Patel, Edmund D. Blackshear

IBM Integrated Supply Chain, Hopewell Junction, New York 12533, USA

Server system requirement drives the evolution of 3D DRAM technologies. The new generation provides form factor, electrical and power performance benefits. Meanwhile it brings in a more complicated design, new assembly technologies and new failure mechanisms. The optimum 3D DRAM technology is determined by complex tradeoff of benefits, cost, time to market and reliability. The characteristics of the optimum 3D DRAM technology for DDR2, DDR3 and future server memory system are analyzed in detail. TSV (Through Silicon Via) DRAM array stacking is expected to bring superior power performance which could be a significant market diver. The tradeoff of this future technology and process development trend is presented.

Server System Requirement Drives 3D DRAM Technology

The server system memory capacity is growing at minimum 2X per generation. System volumetric space allocation limits the total number of memory socket/module and hence memory module density growth is required. In recent years the slowdown of the "next generation die shrink" and cost crossover create a gap between the module density growth requirement and DRAM die density growth. This gap is widening [figure 1] and creates an application space for 3D DRAM technologies.

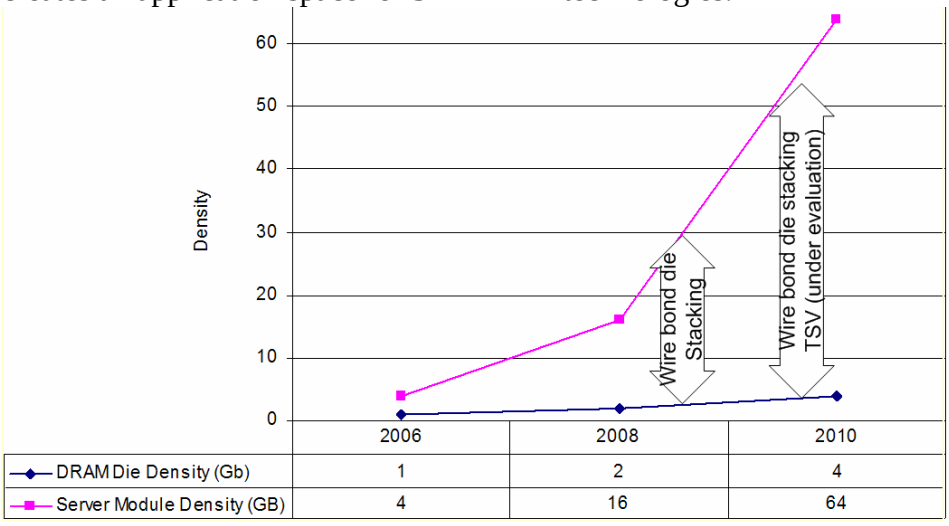


Figure 1. Server memory module density requirement vs. DRAM die shrink

DDR2 3D Technology

DDR2 memory was introduced with BGA monolithic package. There are two competing DDR2 3D technologies - BGA stacking (package on package) and wire bond die stacking. Both technologies were introduced in IBM server platforms around 2005.

Ex. 1016, Page 729

BGA Stacking (Package on Package)

Several BGA stacking designs have been used in IBM server platforms [figure 2]. Those designs share one common feature – individual DRAM die is packaged. In the absence of die level burn in capability, the simple package enables burn in prior to the stacking process, a very important benefit to maintain stack package yield [1]. When the die yield of a new generation DRAM is still at ramp up stage, lacking KGD (Known Good Die) is an issue for stack package yield. BGA stacking offers a workaround of this issue. By 2004, several assembly subcontractors had already developed BGA stacking capability. By using 3rd party BGA stacking service, memory suppliers obtained the "time to market" edge. Yield and time to market advantages were the primary market drivers of BGA stacking technology.

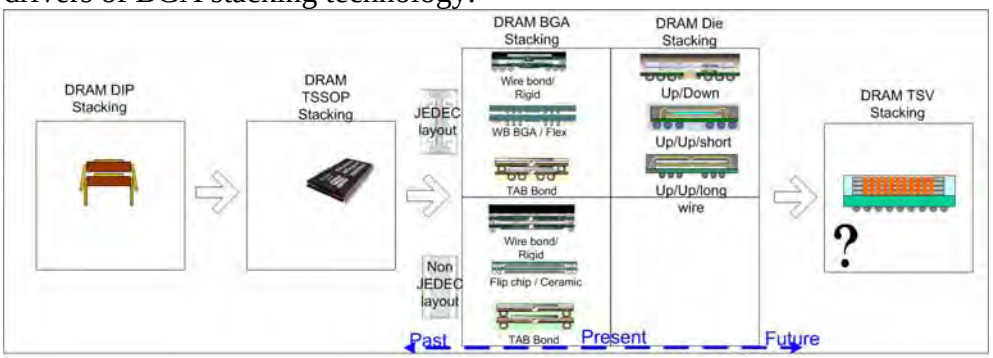


Figure 2. 3D DRAM evolution

Low Profile is a Significant Market Driver

Thermal/cooling enhancement is common for high density DDR2 memory modules [figure 3]. Heat sink and heat spreader consume z space and add impedance to air flow. To permit z space allocation for the thermal enhancement and permit adequate air flow, the low profile 3D DRAM stack packages are required. Thermal management/cooling is critical to DDR2 server memory sub system design therefore height advantage is a significant market driver.

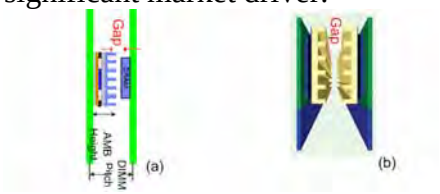


Figure 3. Memory module thermal enhancement schematics a. fin type AMB heat sink b. fin type full module heat spreader

VLP (Very Low Profile) memory module outline was introduced primary for blade server chassis designs. VLP module outline reduces the board real estate for component placement and routing. Therefore 3D DRAM stack packages with small outline (x, y) are desired for high density VLP memory module applications.

BGA Stacking vs. Die Stacking

BGA stacking technologies pay the form factor penalty for the burn in/yield advantage. Some BGA stacking designs achieve JEDEC pin out with additional z height penalty [figure 2]. In contrast with this approach of stacking package on package, die

Ex. 1016, Page 730

stacking designs stack individual die with intrinsic advantage to be thinner/smaller. Establishing yield clip level to allow only high yield DRAM wafers for DDP (Dual Die Package) and QDP (Qual Die Package) is the typical scheme to control stack package yield, which has been proven to be effective for DDP. However QDP yield is still an issue. To illustrate the impact of die yield on QDP yield, formula [1] is under the assumption that defective DRAM die is the only yield detractor for QDP. The yield issue has been the limiting factor for DDR2 QDP to be a low volume premium technology.

$$\text{QDP yield} = (\text{Die yield})^4 = (\text{DDP yield})^2 \qquad [1]$$

DDR2 Die Stacking Technologies

Memory suppliers developed various wire bond die stacking designs. Based on the die orientation, those designs can be classified as up/down or up/up. The up/down DDP grew out of the BOC (Board on Chip) BGA construction and fully leveraged the proven BOC assembly process technology elements: the bottom die assembly is almost identical to BOC. The primary process development challenge was the treatment of long bond wire for the top die. By optimizing wire bonding and encapsulation processes, memory suppliers developed adequate process capability to control the top die long wire sweep. Up/down DDP stacking possesses characteristics of an optimum 3D technology such as small form factor, acceptable stack package yield, relatively short development cycle, near zero capital requirement, low stacking assembly cost adder, and proven stack package reliability. However, up/down construction can't be extended to support 4-height DRAM stacking. In addition, the up/down DDP has intrinsic electrical issue. The electrical path of up and bottom die are asymmetric. 800Mbps could be the data rate support limit for up/down DDP.

Up/up DRAM stacking designs could be divided into two families – short wire and long wire [figure 2] based on the wire bond scheme. The short wire construction requires additional material layer (RDL: Re Distribution Layer) at BEOL to fan out the wire bond pads to die edge. For short wire construction, the treatment of bond wire to spacer interaction was relatively straightforward, an important benefit for several memory suppliers to adopt this construction. This paper will focus on up/up/short wire. For simplification, in the later section for this paper, up/up refers to up/up/short wire. Up/up DRAM stack package development was challenging. Extensive FEM (Finite Element Module) and DoE (Design Of Experiment) were required for geometric design, spacer/die attach material selection and assembly process development. The process optimizations for RDL, wafer dicing, die attachment, low profile wire bond, and encapsulation were critical to stack package quality/reliability performance. While requiring substantial engineering resource, up/up DRAM stack package didn't entail significant capital investment. Some DRAM suppliers were able to extend existing in house manufacturing tools and capacity to manufacture up/up DDR2 stack packages in volume. We believe the up/up D(Q)DP is by far the most complex DRAM 3D technology used in server memory systems. Design for reliability activities was important to catch design oversights like the resin separation of epoxy die attach/spacer material. Under assembly stress and power on/off induced thermal-mechanical stress, the package die interaction could cause the dicing micro crack to propagate. Those issues were adequately addressed by design optimizations and process improvements such as

Ex. 1016, Page 731

silicon spacer, FOW (Film Over Wire), and die attach/encapsulation material optimization to decouple the stress on bottom die.

3D DDR2 Tradeoff Matrix

The tradeoff factors for the DDR2 stack package are summarized in figure 4. Those factors that are critical to server application are highlighted. Die stacking has intrinsic benefit in form factor which is a significant market driver. With the improvement on quality/reliability, die stacking gradually surpassed BGA stacking to be the prevailing technology in server application space.

Trade off factor		BGA stacking	Wire bond Die		Note
Description	Weight		up/down	up/up/ short wire	
Form factor: height (z)	very high	-	+	+	permit air flow and thermal enhancement
Form factor: size (x,y)	medium	-	+	+	critical for VLP
High speed electrical Performance	medium	-	-	+	up/down is proven for 667MHz; might be able to support higher speed with design optimization
Power Performance	medium	n/a	n/a	n/a	
Quality and reliability	very high	-/=	++	- → +	
support 4 height	medium	-/=	-	+	critical for low profile/high density niche applications
Capital Investment for tooling	high	+	+	+	
burn in/KGD yield issue	high	+	-	-	
Assembly process complexity and cost adder	high	=	+	=	
time to market	very high	+	+	-	

Figure 4. DDR2 3D technology evaluation for server application

DDR3 and future DRAM 3D Technology

DDR3 memory was introduced to server applications in 2008. The DDR3 date rate growth creates uncertainty for the existing DRAM stack technologies. The up/down DDP has intrinsic drawback to support high date rate. Very few suppliers in the memory community are optimistic about extending up/down DDP to support 1333Mbps while most focusing development effort on up/up DDP. Development activities on RDL layout/materials optimization and bump/wire bond hybrid inner package interconnection are carried out by the memory industry. We believe advanced RDL and new interconnection scheme will extend up/up DDP to support 1333Mbps and potentially 1600Mbps.

TSV DRAM Array Stacking Concept

A TSV stack package design could completely eliminate bond wire by using solder bump for stack to substrate bonding. By minimizing (or eliminating) bond wire, TSV stacking technology is expected to make DRAM stack packages not only smaller but also capable to support higher data rate.

DRAM array stacking concept has been published by Advanced Device Development Division, NEC Electronics, Elpida Memory and Oki Electric Industry [3]. This concept is illustrated in figure 5. Four (or more) DRAM core chips are stacked via TSV and bonded to substrate together with a separate interface chip of peripheral circuits [figure 5 (a)]. The interface chip could be integrated with bottom DRAM core chip to form

Ex. 1016, Page 732

master-slave architecture [figure 5 (b)]. Both architectures integrate DRAM peripheral circuits, reduce I/O buffers and registers and hence achieve power reduction.

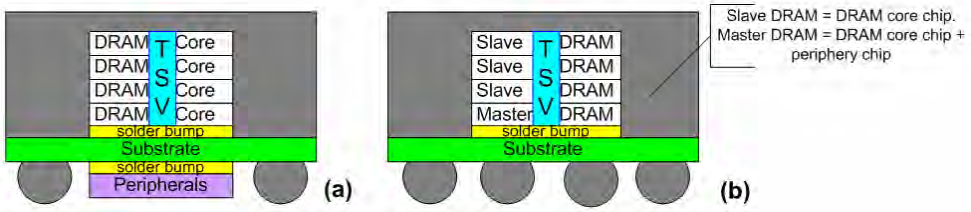


Figure 5. Conceptual schematic of DRAM array stacking

Superior Power Performance Could be a Significant Market Driver

Several DRAM suppliers have developed FOW or silicon spacer based thin D(Q)DP technologies. We expect wire bond based die stacking technology to support DDP with 1.2mm maximum height and QDP with 1.5mm maximum height. Those projected stack package height is adequate for the most complicated DDR3 memory thermal enhancement/cooling structure under development. Therefore form factor benefit alone will not provide adequate justification for TSV DRAM.

The power/thermal issues dominate server memory system design approach. TSV DRAM array stacking technology's power saving benefit is very important for DDR3 and future server memory system design. For high density high speed server memory module, the superior power performance could be a significant market driver for TSV array stacking technology.

TSV DRAM Array Stacking Development Challenges

For TSV DRAM array stacking, memory suppliers have little production experience and capability for the critical enabling technology elements, such as via forming/filling, wafer thinning, bump forming, and die to die/substrate bonding. To transform TSV DRAM array stacking from concept to production, the capital investment in tooling and significant development effort are required. To satisfy server reliability requirement, a TSV DRAM package need to demonstrate the integrity of critical structures such as through silicon via, micro bump, die to die/substrate bond, underfill layers and RDL. The treatment of die to package interaction is critical to TSV DRAM package reliability since via forming and metal filling will change the mechanical property of DRAM dies. Treatment of micro bump bonding and underfill is critical to minimize pressure point on the active layer of DRAM die. The VRT (Variable retention time) issue induced by thin DRAM die is another reliability concern needs to be addressed [2]. DRAM TSV stack yield will affect the assembly cost adder for each package. Electrical repair or other yield improvement schemes need to be developed to control the assembly cost.

DDR3 and Future DRAM 3D Technology Tradeoff Matrix

For DDR3 (and future DRAM) 3D technologies, the critical tradeoff factors for server memory system are highlighted in figure 6. Power saving benefit is the most important driver for server memory system designers to push memory suppliers for TSV array stacking. However the development requires substantial capital investment and

Ex. 1016, Page 733

new assembly technology development. New failure mechanisms need to be contained at development stage.

Trade off factor		TSV array stacking	Wire bond Die stacking*	Note
Description	Weight			
Form factor: height (z)	very high	+	+	thin wire bond die stacking could support 1.2mm DDP & 1.5mm QDP
Form factor: size (x,y)	medium	+	-	critical for VLP
High speed electrical Performance	high	+	=	Several wire bond die stacking designs could support 1333MHz; supportability to 1600MHz TBD
Power Performance	extremely high	++	-	
Quality and reliability	very high	-	+	
support 4 height or more	medium	+	=	critical to low profile/high density niche applications
Capital Investment for tooling & volume production capacity	very high	-	+	
burn in/KGD yield issue	high	-	=	
Assembly process complexity	high	-	+	
time to market	very high	-	+	

*: up/up/short wire

Figure 6. DDR3 (and future DRAM) 3D technology evaluation for server application

TSV DRAM Array Stacking Assembly Process Development Trend

To minimize DRAM device architecture modification, minimize exposure to elevated assembly temperature, and utilize the proven assembly technologies and existing tooling where possible are the typical mindsets for early TSV DRAM array stacking development. The assembly process development trend is summarized in figure 7. To minimize or avoid DRAM architecture modification, the process sequence for early product will have via formation process after FEOL and prior to the last metal process (or even after last metal process). Direct Cu bonding will not be selected for early product due to the concern of high assembly temperature induced compromise on DRAM device characteristics. To shorten the process development cycle, the stack method for early product might be chip to chip (C2C) and some memory suppliers might select wire bonding for stack to substrate interconnection.

	via first		via last	
Process sequence	before FEOL	after FEOL	before Thinning/bonding	after bonding
Via formation		plasma etching	plasma etching	
Via filling (metallization)		Cu	Cu	
Wafer thinning		before bonding	before bonding	
Bonding		solder micro bump; CPB	solder micro bump; CPB	
Stack method		C2C	C2C	
Interconnection to substrate		wire bond; flip chip	wire bond; flip chip	

Figure 7. Projection of TSV DRAM array stacking assembly technology for early product

REFERENCES

1. E. Blackshear, 2 *Terabyte memory system packaging*, IEEE system packaging workshop, (2004).
2. W. Bornstein; R. Dunn; T. Spielberg, *Field degradation of memory components from hot carriers*, 44th annual International Reliability Physics Symposium, (2006).
3. Kawano, M.; Uchiyama, S.; Egawa, Y.; Takahashi, N.; Kurita, Y.; Soejima, K.; Komuro, M.; Matsui, S.; Shibata, K.; Yamada, J.; Ishino, M.; Ikeda, H.; Saeki, Y.; Kato, O.; Kikuchi, H.; Mitsuhashi, T., *A 3D Packaging Technology for 4 Gbit Stacked DRAM with 3 Gbps Data Transfer*, (2006).

Ex. 1016, Page 734

ECS Transactions, 18 (1) 707-712 (2009)
10.1149/1.3096525 © The Electrochemical Society

3D Integrated, Ultra-Thin Functional Microcontroller Device for Wireless, Flexible ECG Systems

W. Christiaens[a], T. Torfs[b], W. Huwel[c], C. Van Hoof[b], and J. Vanfleteren[a]

[a] ELIS Department, Ghent University, Belgium
[b], IMEC, Leuven, Belgium
[c] ACB, Dendermonde, Belgium

This paper presents a technology for the 3D-integration of a TI MSP430 low-power microcontroller inside conventional double sided flex substrates, for a wearable wireless ECG system. The devices are first thinned down to 25 µm and next integrated as a UTCP (Ultra-Thin Chip Package) in between two polyimide layers, resulting in 50-60 µm thin chip package.

These extremely thin, mechanically flexible packages serve as interposers for embedding into conventional multilayer flex substrate, replacing for example the integration of naked dies. Providing a contact fan-out, this approach enables the testing of the device before embedding (solving the Known Good Die issue). It also eliminates the need for precise placement and the use of expensive ultra high density printed circuit boards.

Smaller SMD components have been mounted above and below the embedded chip, leading to a high density of integration. In addition, the embedded thin silicon chip becomes mechanically flexible itself, leading to an increased flexibility of the resulting system.

Introduction

One of the main challenges in the electronics manufacturing and packaging development is how to integrate more functions inside the same, or even smaller, size. The electrical performance and the number of functions of every new product generation are increasing while the size and the weight of the products are decreasing. To meet this, the semiconductor industry is integrating more and more transistors into the same silicon area. This allows either to reduce the size of the IC or to increase the functionality and performance of the same size IC components. At the same time there is also an interest to increase the packaging density. While silicon chips continue integrating more functionality as per Moore's law, the packaging is challenged to also integrate and shrink.

A high degree of miniaturization can be achieved by removing the assembled active devices from the surface and integrate them inside the inner layers of multilayer FCB or PCB boards.

On flex substrates typically only pure electronic assembly is performed up to flip-chip components. Embedding of passive or active components in conventional flexible printed circuits is at least not state of the art. This limitation to two component layers (front and backside of the flex laminate) limits of course the compactness of the circuit, and the presence of relatively large rigid components also limits the flexibility of the FCB.

Ex. 1016, Page 735

This paper now presents the embedding of an ultra-thin, functional microcontroller device inside a double sided flex board, by means of a unique concept for packaging ultra-thin chips: the Ultra-Thin Chip Package (UTCP).

The UTCP serves as flex interposer and can be used for the embedding inside the substrate, replacing for example the direct integration of bare dies. The UTCP allows for easy testing of the chip before embedding, solving the KGD issue, and provides a contact fan out with more relaxed pitches, eliminating the need for very fine pitch PCB or FPC compatible with the chip contact pad pitch.

The 3D integration of UTCP packages leads to high density integration, since SMD components can be mounted on top and bottom of the integrated devices. In addition, the UTCP packaged thin silicon devices are mechanically flexible itself, leading to an increased total flexibility of the resulting system.

Wireless ECG system

The demonstrator consists of a high-density flexible ExG circuit for personal health or wellness applications. IMEC's ExG system is a bio-potential wireless sensor node, able to monitor the vital body signs provided by portable electrocardiogram (ECG, which monitors the heart activity), electromyogram (EMG, which monitors muscle contraction) and electroencephalogram (EEG, which monitors brain waves). The system collects and processes data from (external) human body sensors and wirelessly transmits the data to a central monitoring system. Small size and low power consumption of the system enables non-invasive and ambulatory monitoring of vital body parameters.

The block scheme of e.g. an EMG circuit is given in Figure 1. The system uses IMEC's proprietary ultra-low-power bio-potential readout ASIC (application specific integrated circuit) to extract the bio-potential signals produced during the ECG, EMG or EEG measurements. The low power microcontroller MSP430F149 drives the bio-potential chip, also digitalizes the samples with its A/D converter. The signals are finally sent by the 2.4GHz low power radio Nordic NRF2401A using a coplanar antenna. Information can be transmitted to a pc by a 'USB stick' receiver for visualization and recording.

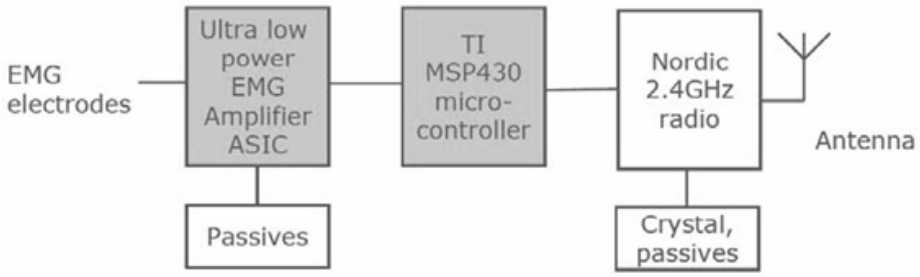


Figure 1. Block scheme of an EMG circuit.

This ECG system is realized on standard double sided flex laminates. Such flex substrates ensure a certain degree of mechanical flexibility for the wearable systems. However, an important factor limiting miniaturization and flexibility is typically the presence of relatively large rigid components. One chip, the TI microcontroller device, is

Ex. 1016, Page 736

now removed from the surface and embedded as UTCP inside a commercial standard flex substrate. Smaller SMD components can be mounted above and below the embedded chip. This leads to a high density of integration. In addition, the embedded ultra-thin silicon chip becomes mechanically flexible itself. This leads to an increase in the flexibility of the resulting system.

UTCP packaging of ultra-thin, functional microcontroller

The microcontroller used for the wireless ECG application is the Texas Instruments MSP430F149 low-power microcontroller with integrated 12-bit analog-to-digital converter.

The TI microcontroller was only available as unthinned, bare, single dies. The individual dies were first thinned down individually to only 20-25 µm, by a combined lap- and polishing process.

UTCP package production

After thinning, the microcontroller devices are first integrated as a UTCP package. An overview of the individual process steps of the Ultra-Thin Chip Package production is depicted in Figure 2. The base substrates are a 20 µm polyimide layer spincoated on a rigid glass carrier. For the fixation and the placement of the chips a benzocyclobutane (BCB) is used as adhesive. The chip is covered with a next 20 µm thick polyimide layer. For the contacting to the chip, contact openings to the bumps of the chips are laser drilled and a 1 µm TiW/Cu layer is sputtered and photolithographically patterned. This metal layer provides a fan out to the contacts of the chips.

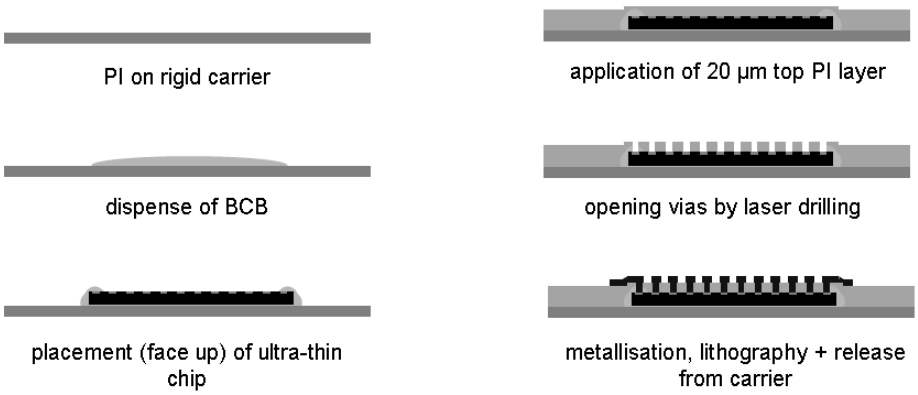


Figure 2. UTCP process flow.

After processing, the whole package can be released easily from the rigid carrier. The result a is a very thin, even flexible, chip package, with a total package thickness of only 50–60 µm.

Ex. 1016, Page 737

Design

The top metallization on the UTCP package provides a fan-out of the smaller pitch contacts of the chip. The metallization is shown in Figure 3: it includes a small fan-out of the 64 die contacts to 400 x 400 μm^2 contacts with pitch of 500 µm, and a larger fan-out to 650 x 1300 µm contacts. The outer, large contacts can be used to easily connect the integrated microcontroller for programming and testing. The inner contacts match the design rules of conventional flex board manufacturers and will be used for embedding the UTCP package in multilayer flex boards. Outer dimensions are 9 x 9 mm^2 for the small fan-out and 3 x 3cm^2 for the large fan-out. The design also includes solder pads to solder mount an SMD resistor and a LED component on top of the package for visual functionality demonstration after processing.

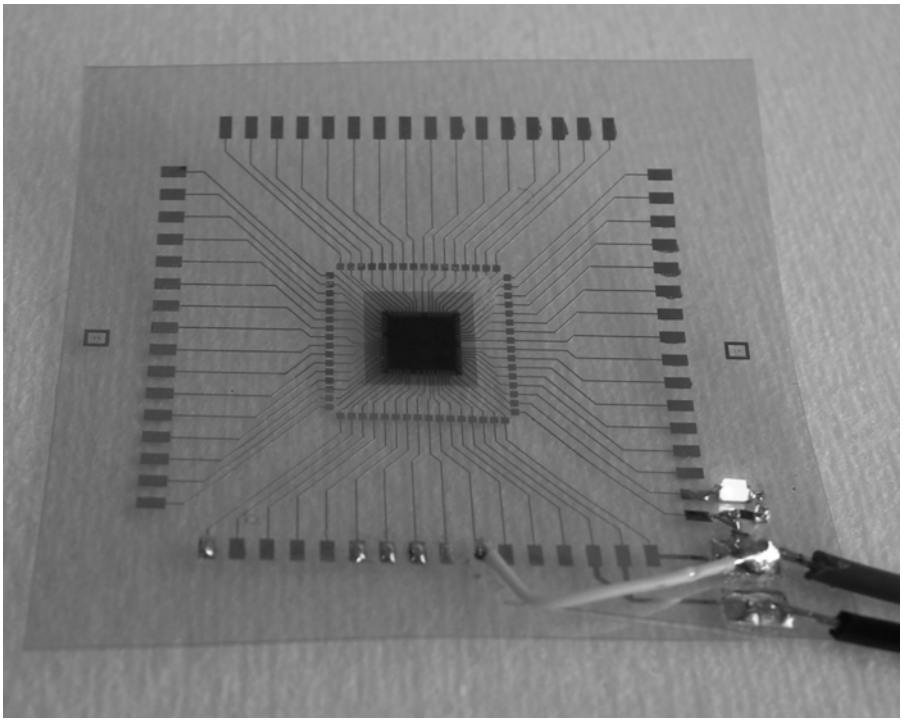


Figure 3. UTCP package of MSP430F149 microcontroller.

Functionality testing

The fan-out metallization of the UTCP package allows for easy testing of the thinned microcontroller devices, after processing and before the integration inside the flex substrates. First the microcontroller was programmed to generate a blinking sequence on the LED to demonstrate its functionality: see Figure 3. But also more extensive tests of the analog-to-digital converter of the embedded microcontrollers are performed after UTCP integration. The linearity and also the AC specs (including signal to noise ratio (SNR), effective number of bits (ENOB), total harmonic distortion, spurious free dynamic range) were compared, between a bare, unthinned microcontroller die and UTCP packaged thinned microcontroller. The UTCP performs slightly worse than the naked die, but the small deviations are limited and will have only little effect on the usability of the packaged microcontroller.

Ex. 1016, Page 738

Flex circuit production

After testing, the UTCP package is embedded in turn inside a double-layer flex printed circuit board (PCB) using standard flex PCB production techniques. This was done at ACB, a Belgian flex manufacturer.

The principle is as follows: the UTCP packages are aligned and fixed (heat tack) on a patterned inner layer of the multilayer substrate. Before the mounting of the package, first an adhesive sheet is prelaminated on the patterned inner layer. The packages are fixed by heat tack on this panel, before the multilayer stack is built-up and laminated together using exactly the same process as for conventional multilayer board production.

A first flexible (double sided) laminate is patterned at only one side (using photolithography and etching) and an adhesive sheet is laminated on top of the patterned side. This pattern contains also a copy of the contact pad layout as on the UTCP fan out. Next, the UTCP package is aligned manually with it contacts on the alignment contacts of the panel and it is fixed by heat tack (using a solder iron). Once the package(s) are fixed on the adhesive top layer, a single sided laminate is laminated on top of these packages. Next, the through holes are drilled, for the interconnection between both sides of the panel, but also for the interconnections to the UTCP contact pads. Finally the through holes are metallized, as for a standard multilayer flex.

After substrate production, components are assembled on the flex board. The result after assembly is shown in Figure 4: the microcontroller device is removed from the surface and integrated inside a standard flex substrate. Small SMD components are even mounted on top and/or bottom of the integrated component.

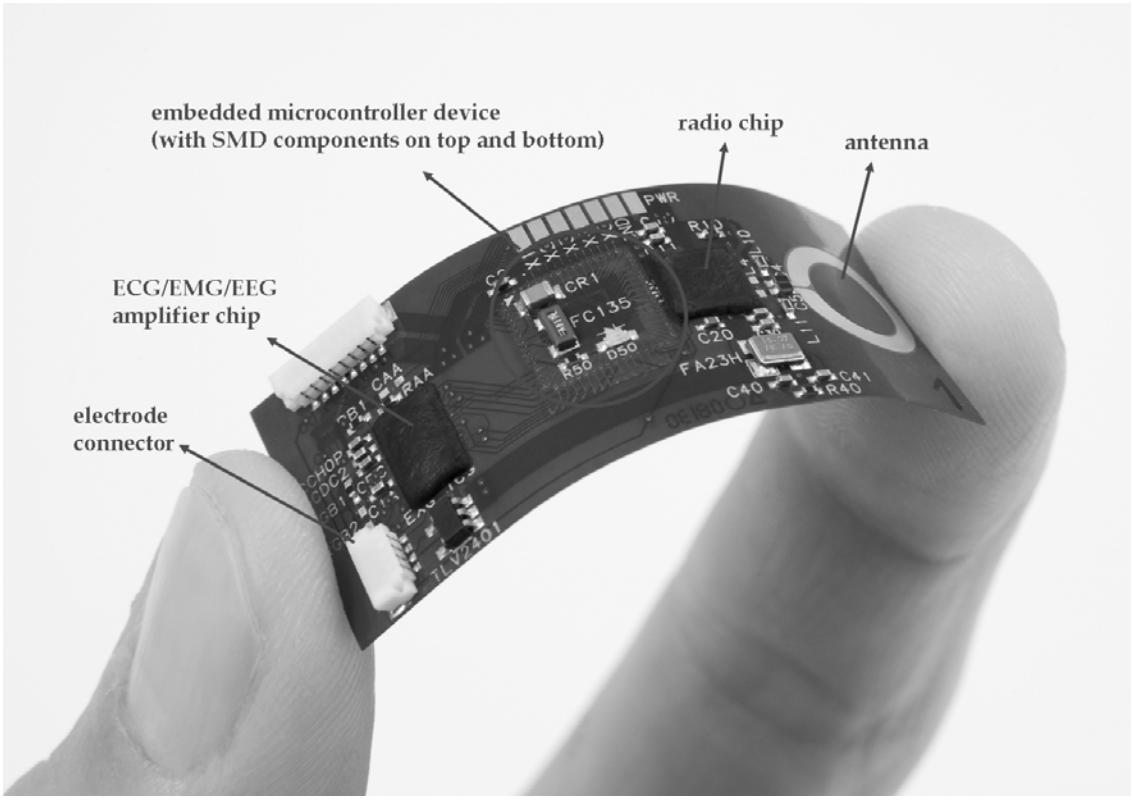


Figure 4. The flexible wireless biopotential system and its components (excluding battery).

Measurements

The functionality of the wireless biopotential system was tested after assembly. The final ExG demonstrator with integrated microcontroller device is completely functional, and able to monitor ECG signals. Figure 5 shows an example ECG waveform recorded with the wireless system using standard disposable stick-on Ag/AgCl foam electrodes.

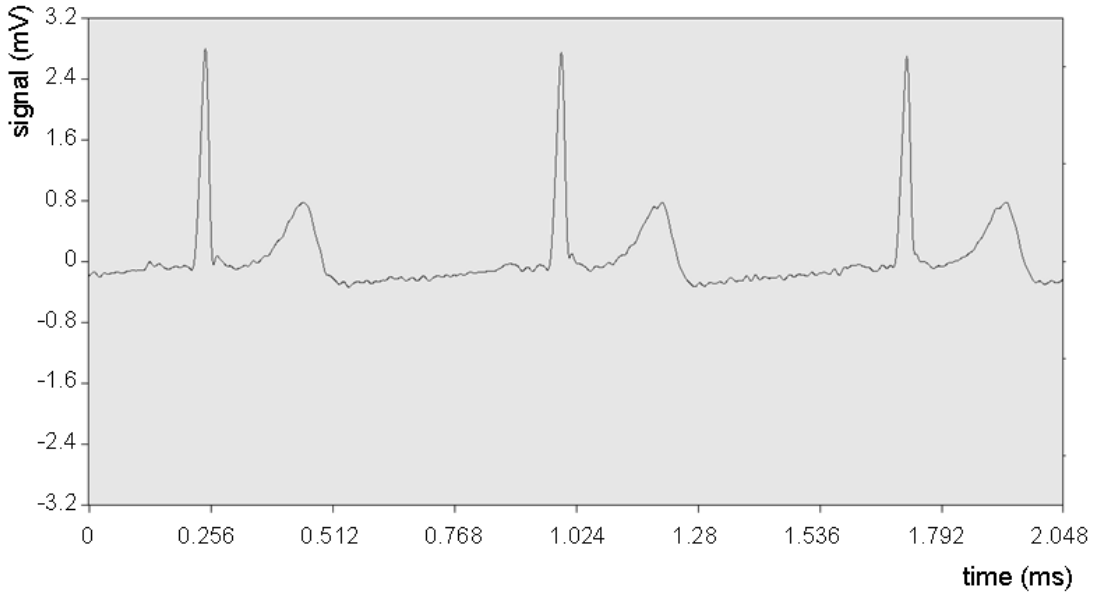


Figure 5. The flexible wireless biopotential system and its components (excluding battery).

Conclusions

A wireless ECG system, which uses a microcontroller device integrated inside the substrate, is presented in this paper. This 3D integration of this relatively large components, not only leads to an increased miniaturization, but also enhances the mechanical flexibility of the wearable system.

The other active components in the system are an ultra-low power biopotential amplifier chip developed by IMEC and a commercial off-the-shelf 2.4GHz low power radio transceiver (Nordic nRF2401A). These components can in principle be embedded inside the flex PCB just like the microcontroller. This will increase the flexibility and integration density of the device even more.

Acknowledgments

The work was carried out under the EC IST funded project IP-SHIFT (contract number 507745).

Ex. 1016, Page 740

ECS Transactions, 18 (1) 713-719 (2009)
10.1149/1.3096526 © The Electrochemical Society

Challenges, Trends and Solutions for 3D Interconnects in Lithography and Wafer Level Bonding Techniques

Margarete Zoberbier[a], Erwin Hell[b], Kathy Cook[c], Marc Hennemayer[a], Dr.-Ing. Barbara Neubert[a]

[a] SUSS MicroTec Lithography GmbH, 85276 Garching, Schleissheimerstr. 90, Germany
[b] SUSS MicroTec Shanghai Co., Ltd. Room 3703 Nanzheng Building, 580 Nanjing Road West, Shanghai 200041, P. R. China
[c] SUSS MicroTec Inc., 228 Suss Drive, Waterbury Center, VT. 05677, USA

Technology advances such as 3-D Integration are expanding the potential applications of products into mass markets such as consumer electronics. These new technologies are also pushing the envelope of what's currently possible for many production processes, including lithography processes and wafer bonding.
There is still the need to coat, pattern and etch structures. This paper will explore some of the lithographic challenges associated with 3D interconnection technology. Wafer bonding techniques as used in the 3D Packaging will be described with all the challenges and available solutions and trends. Furthermore a new Maskalinger technology will be introduced which allows extreme alignment accuracy assisted by pattern recognition down to 0.25µm.
An overall introduction on the challenges, trends and solutions for 3d interconnects in lithography and Wafer Level bonding techniques and the SUSS's equipment platform will be described accordingly to the needed processes. The processing issues encountered in those techniques will be discussed with a focus on wafer bonding and lithography steps.

Introduction

The expanding consumer electronics market is clearly driving the development of today's semiconductor innovation. The push for integration, reduction in power consumption and the need for smaller form factors lead to new architectures which combine dissimilar technologies and lead to creative packaging methods, where maximum functionality is packaged into minimal space. So 3D Integration is considered the next generation packaging solution.

Today different 3D packaging approaches like SiP (System in Package), SoC (System on Chip) and SoP (System on Package) have been developed in order to answer the requirements for smaller footprint, shorter interconnects and higher performance.

SiP "System in a Package" is a functional system or subsystem with multiple wire-bonded or flip-chip dies in an IC package. Other components are placed on the motherboard, like passives, SAW/BAW filters, pre-packaged ICs, connectors and micromechanical parts. This technology enables a stacked chip package with reduced form factor.

SoC (System on Chip) integrates all the different functional blocks, like processor, embedded memory, logic core and analog in a monolithic way. These blocks are required to integrate the system design on a single semiconductor chip. SoC designs usually consume less power and have a lower cost and higher reliability than the multi-chip

Ex. 1016, Page 741

systems that they replace. And with fewer packages in the system, assembly costs are reduced as well.

SoP (System on Package) uses through-vias and high density wiring in order to achieve a higher miniaturization. It is an emerging microelectronics technology that places an entire system on a single chip-size package. Where “systems” used to be bulky boxes housing hundreds of components, SoP saves interconnection time and heat generation by enabling a full system with computing, communications, and consumer functions all in a single chip [1] [2].

Through Silicon Via (TSV) has evolved as one of the key technologies for 3D integration and wafer level packaging. 3D TSV has the potential to one day replace wire bonding and thus enable further size and cost reduction, which is one of the biggest challenges, and increase the performance of the device.

Today 3D TSV technology has become critical to the growth of 3D components integration, like memory stacking, or for MEMS structure packaging. The first application which is using the TSV as mainstream technology is the packaging of CMOS image sensors (CIS). For CMOS image sensors, WLP is already an industrial reality. Today, already about 35% of CMOS imager sensors can be found into latest consumer cell-phones and notebook cameras are encapsulated in a WL-CSP and this number keeps growing (Fig. 1.) [3] [4].

Process Flow

One of the typical process flows to form TSVs is shown in Figure 2. These steps are required for through silicon via wafer processing. First, the etch mask must be created. This involves coating, exposing and developing the mask. Once the mask is created, the vias can be etched and insulated. Via filling can then be completed using various materials such as copper and tungsten. The fill process is determined by the fill materials. As of today, copper is the most commonly used material for TSVs, but other materials like Tungsten (W) or Cu3Sn alloy are used as well.

Process Details

Exposure and Development

Photolithography of via openings in photo resist looks pretty straight forward. However, the various following process steps and different via sizes require specific photo resist exposure and development conditions and ask for an optimized set of parameters.

Typically via sizes down to 5µm (Fig. 3) can be easily and cost effectively achieved by 1X full field photolithography. State-of-the-art resolution limits are at about 3µm in proximity printing on 300mm substrates. However, tight CD control of via openings requires accurate gap setting, excellent light uniformity and exposure dose control. All this will affect the final exposure results and therefore need to be accurately controlled.

For the described experiment a typical via test mask having different via diameters was used. The SUSS MA300 Gen2 Mask Aligner provides an average intensity of about 90mW/cm² (broadband) coupled with a light uniformity of less than 3% over the whole 300mm wafer. All wafers were exposed with an exposure gap of 20µm. Figure 3 shows exposed and developed vias with a diameter of 3µm in AZ1505 and further examples.

Ex. 1016, Page 742

The development process was again carried out on the ACS300 Gen2 using an aqueous develop module that was equipped with a binary spray dispense system. The binary spray nozzle setup and water-jacketed dispense line with temperature control to point-of-use allow for reducing process times and minimizing material consumption. Variable rate arm movement across the wafer was employed in order to optimize the uniformity of the spray process. Both, AZ4110 and AZ9260 were developed using a 1:4 dilution of AZ400K and deionized water. TMAH based AZ726MIF was used for the development of AZ1505.

The trend for smaller vias also requires accurate overlay of the print result. Overall alignment performance of the mask aligner as well as runout effects are the two main important factors that influence the overlay result. The SUSS MA300 Gen2 uses two novel technologies to achieve highly accurate overlay.

Alignment accuracies below 0.5µm (3sigma) can be achieved on the MA300 Gen2 by adopting the DirectAlign® technology. First of all, the system accurately aligns wafer and mask in alignment gap, which is typically set to 50 - 100µm. The final high accuracy alignment step is performed in exposure gap. Advanced pattern recognition software measures and controls the alignment, thus achieving sub-micron alignment accuracy.

Also alignment accuracies down to 0,25µm can be achieved now using the MA/BA8 Gen3 Maskaligner. The new Operator-assisted alignment technology with continuously automatic calculation of achieved alignment (X, Y, Θ) between both targets allows extreme alignment accuracy assisted by pattern recognition. During manual alignment the COGNEX® based pattern recognition software continuously measures the achieved accuracy and reports it to the operator. With its sub pixel resolution the system supports highest alignment precision, prevents misalignment and maximizes yield. So the need for high alignment accuracy is being answered with this technology.

Besides an excellent equipment performance in terms of alignment accuracy, the control of mask and wafer temperature is also crucial to achieve optimum overlay results on 300mm wafers. The MA300 Gen2 employs a temperature controlled exposure chuck to keep a constant and uniform wafer temperature and to compensate for runout effects. In a proximity printing system, mask and wafer are in close proximity to each other, assuring that the mask temperature is indirectly controlled by the chuck. Any runout, measured via pattern recognition, can be reduced by changing the temperature in the exposure chuck (Fig. 4). The ThermAlign® technology is able to reduce runout down to 0.2 to 0.3µm.

During the experiment the ThermAlign® chuck was set to 22°C. The alignment was done using an automatic alignment system in direct alignment mode with SUSS proprietary alignment targets.

The achieved overlay data on 300mm wafer are shown in table I and the x-y plot is shown in figure 5.

Bond Alignment

One of the most important topics in 3D Stacking is the alignment accuracy. As can be seen from the roadmap shown in fig. 6, the trend of via sizes is that via diameters will continue to shrink over the next couple of years.

Today via diameter in range of 25 to 75µm are used in CMOS Image Sensors devices. In memory devices typically few microns diameter is used for a TSV. For the wafer

Ex. 1016, Page 743

bonding technology, shrinking via diameters have a direct effect on the post bond alignment accuracy that is required. There is always a minimum overlap requirement so that metal vias have good electrical connection with minimized resistance. This minimum overlap requirement directly translates into post-bond alignment accuracy, which will be in the sub-micron range within the next couple of years.

As both of the wafers which need to be aligned for 3D stacking have metal layers IR alignment becomes impossible. Another approach is the Inter Substrate Alignment. This technique is capable of achieving the required alignment accuracies outlined in figure 6.

In this alignment technique special optics are used that are inserted between the two wafers. The ISA objectives, on left and right, image the alignment key on the upper and lower wafer simultaneously. The alignment stage moves the wafers and a pattern recognition algorithm performs the alignment according to the recipe.

After the objectives are retracted both wafers move into contact or to a predefined gap. Typically moving the wafers in the z-axis can create some misalignment. Therefore new precision optics and mechanics have been combined with global inline calibration methods to achieve sub-micron alignment accuracy in the BA300UHP (Ultra High Precision) Bond Aligner.

Bonding

Wafer bonding techniques used for 3D Stacking are mainly:

- metal to metal diffusion bonding (Cu)
- metal eutectic bonding (Cu/Sn)
- silicon fusion bonding
- adhesive bonding (BCB)

There are a lot of pros and cons for each kind of bonding process. Which one will be chosen depends on the application and its requirements. However, there is a clear trend for metal to metal diffusion bonding using copper as bond layer. The advantage of this bond process is that the electrical as well as mechanical connection is done simultaneously.

Copper-Copper Bonding When two metals are pressed together under applied force and heat, the atoms can migrate from lattice site to lattice site bonding the interface together. Such diffusion processes require intimate contact between the surfaces since the atoms move lattice vibration. Copper or Aluminum are optimal for such diffusion processes due to their ductile properties and fast diffusion rates. Copper bonding requires temperatures in the range of 300°C up to 400°C to achieve a good hermetically seal interface. The oxidation of copper can hamper the bonding process. It can be cracked by applying a high force during the bonding or to incorporate a vapor cleaning process to remove the surface oxidation. Using the vapor from formic acid (HC00H) in a bubbler, SUSS MicroTec has developed a point of use removal system for metal oxides. Wafers are placed in cleaning chambers and exposed to the vapor for a few minutes. The chamber is purged and by controlling the partial pressure of O2 in the module no salts are formed that required post process rinses. The process is totally dry and the copper surface is stable for several hours. [5]

Ex. 1016, Page 744

Metal Eutectic Bonding. Eutectic bonding is also of interest for 3D stacking due to the low temperatures, which are beginning at 231°C for Cu3Sn (Fig. 7). Thus the most often requested eutectics are AuSn, AuSi, AlGe, and CuSn, as already mention. It also required the usage of inert gas to prevent oxidation.

Silicon Fusion Bonding. Fusion Bonding is interesting for 3D stacking due to the very short process times and high bond strengths. Nevertheless during the fusion bonding only the mechanical contact is created. However, there are no direct electrical connections across the interface to interconnect the metal layers. Additional processing is therefore required to etch through the wafer stack and backfill vias with metals. This is done by thinning one of the substrates to several tens of micrometers and follow by a patterned etch and metal backfill in a process known as "via last" processing.

The drawback to fusion bonding is the requirements for surface flatness and roughness. Another issue is the limitation in the annealing temperature due to the used metal materials in the wafers. But using plasma treatment the annealing temperature can be reduced from ~1000° down to the required 200°C - 400°C.

Silicon fusion bonding needs the following process steps:

- plasma treatment creating a hydrophilic surface with a specific chemistry and contact angle
- DI water cleaning combined with reactivation and bonding
- temperature annealing in a standard furnace (batch process)

Adhesive Bonding. Adhesive bonding is very often used in 3D due to the low range of process temperature and the topography tolerance. The most often used polymer for precision 3D adhesive bonding is BCB (benzocyclobutene) and is bonded from 150°C-320°C. The flow properties of BCB can be manipulated in order to provide low temperature bonding and alignment accuracies in the range of 1-2µm. BCB needs to be coated on one or both wafers, the use of adhesive promoter AP3000 is absolutely essential. Before bonding the BCB needs to be baked out. If the procuring is below 150°C not all solvents are removed. So the bond will almost certain result in a poor mechanical connection.

Acknowledgments

I would like to acknowledge all the co-authors of this paper for their support in creating this paper.

References

1. Keith Cooper, Kathy Cook, Bill Whitney, *Lithographic challenges and solutions for 3D Interconnects*, IWLPC, Oct. 13-16, 2008, San Jose, CA.
2. Rao R. Tummala, *SOP: What is it and why? A new Microsystems-Integration Technology Paradigm-Moore's Law for system integration of miniaturized convergent systems of the next decade*", IEEE Transactions On advanced Packaging, Vol 27, Nr 2, May 2004, pp 241 -249
3. 3 D. Henry, F. Jacquet, M. Neyret, *Through Silicon Vias Technology for CMOS Image Sensors Packaging*, ECTC, May 27-30, 2008, Lake Buena Vista, FL
4. Eric Mounier, Jérôme Baron, Jean-Christophe Eloy, *A market & technology analysis of WLP solution for IC's, CMOS Image Sensors & MEMS*, ECTC, May 27-30, 2008, Lake Buena Vista, FL

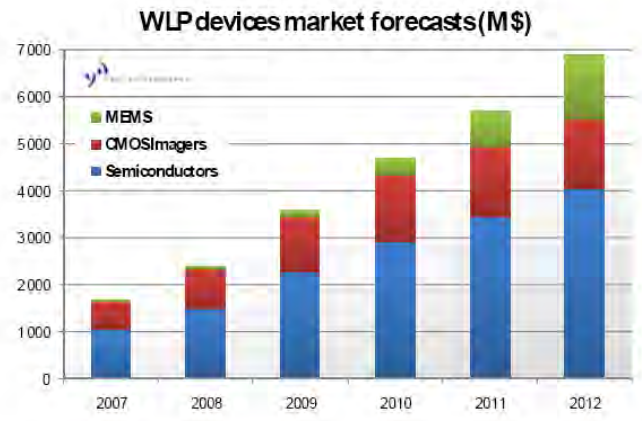


Figure 1. WLP devices market forecasts [4]

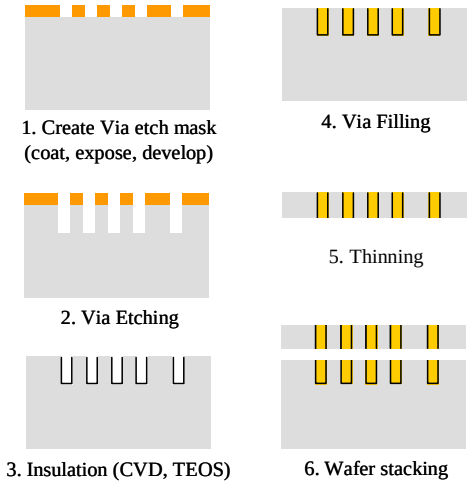


Figure 2. Typical TSV formation flow

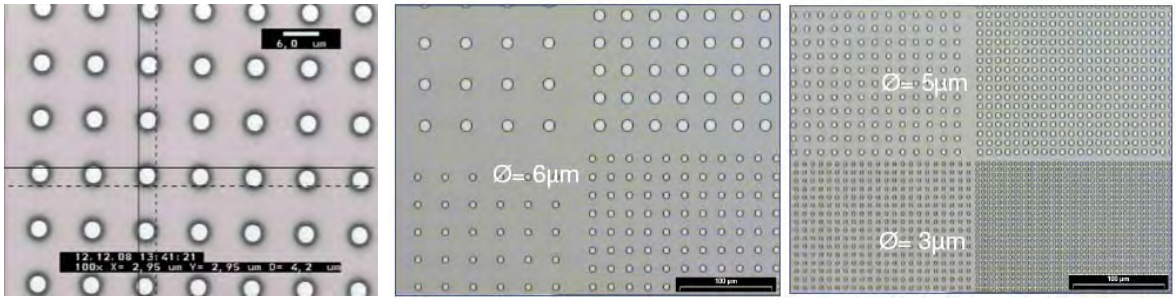


Figure 3. Exposed and developed via with a diameter of 3µm in AZ1505, exposed and developed vias with various sizes (6, 5, 3µm)

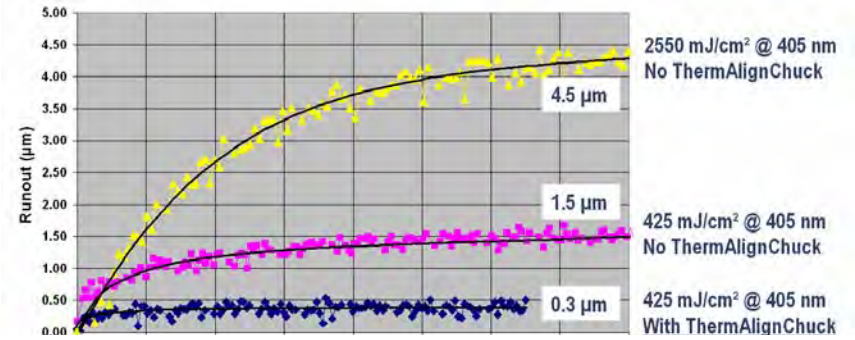


Figure 4. Runout effect between wafer and mask with and without temperature control

TABLE I. Alignment accuracy measurement results.

	Data left side		Data right side	
	x [µm]	y [µm]	x [µm]	y [µm]
Mean	0.02	-0.12	-0.20	0.08
Range/2	0.15	0.20	0.20	0.25
3σ	0.30	0.30	0.32	0.35
Mean+Range/2	0.17	0.32	0.40	0.33
Overlay	0.32	0.41	0.52	0.42

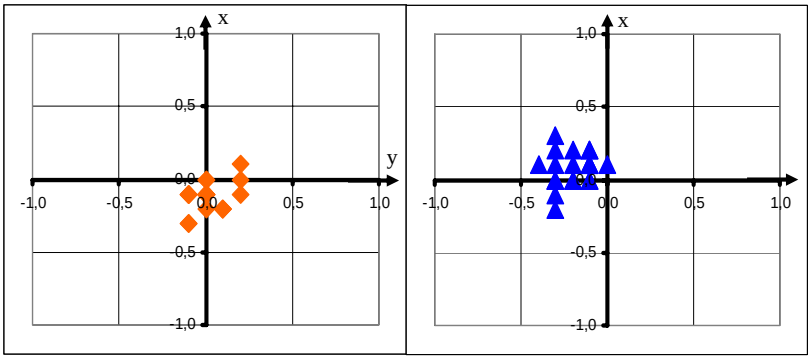


Figure 5. X-Y-plot of the measurement results

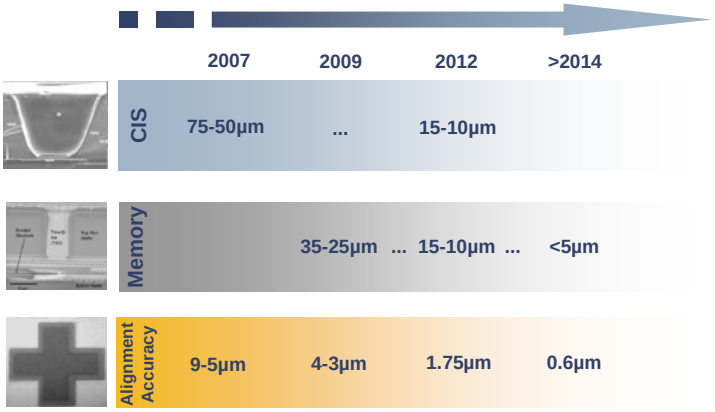


Figure 6. Post-bond Alignment Accuracy Roadmap[4]

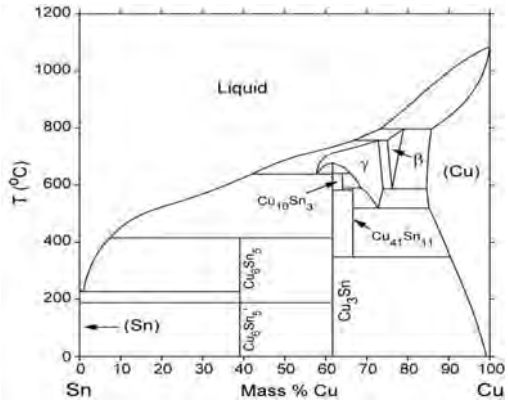


Figure 7. CuSn phase diagramm

ECS Transactions, 18 (1) 721-726 (2009)
10.1149/1.3096527 © The Electrochemical Society

Throughput Enhanced Flip-Chip to Wafer Bonding: The Advanced Chip to Wafer Bonding

A. Sigl, S. Pargfrieder, T. Matthias, M. Wimplinger, P. Kettner

EV Group, DI Erich Thallner Strasse 1, St. Florian/Inn, 4782, Austria

System-in-Package (SiP) is a semiconductor packaging and integration approach to lower costs, increase performance and decrease form factor. Because of the shrinkage of the functional structures on the chips, metal joint technology has to be used, which allow smaller pad sizes for the interconnection of the chips but as drawback needs more time and force to form the connection. Therefore the throughput at standard flip chip bonding drops dramatically.
The Advanced Chip to Wafer bonding process is a two step process for bonding chips on wafer. Chips are aligned and tacked temporarily to a wafer with high throughput and then all chips are bonded in parallel permanently to the wafer with application of force, temperature and under vacuum or process gases. This process flow enables higher throughput and is therefore well suited for bonding of SiPs of high integrated chips in production.

Introduction

In microelectronics there is a continuous demand for smaller devices with higher performance and more functions. One approach is System in Package, the integration of different chips with various functionalities in one package (1, 2). The shrinkage of the electronic devices causes directly the need for smaller package sizes of IC chips. Therefore all important parts of a package have to shrink, for example also the connection of the semiconductor chip with the outside. Actually two main interconnect technologies are used, wire bonding and flip chip (1). Wire bonding is still the workhorse of the semiconductor industry and the dominant method for the interconnection to semiconductor devices. However flip chip gains more and more importance because of its advantages like higher throughput because of the parallel forming of all electrical connections, improved electrical performance due to short connections and higher assembly compactness. There are several approaches for making the electrical connection with flip chip bonding as for example the usage of anisotropic conductive film (ACF), isotropic conductive adhesive (ICA), conductive polymers, gold studs, copper/gold pillars or solder bumps, formed either by solder paste or solder balls. Focused on the solder bump approach, this work shows some effects of joint shrinking of the flip chip process and to suggest an improved process flow, known as Advanced Chip-to-Wafer bonding, suited also for high volume production by providing a high throughput.

Shrinked Controlled-Collapse Chip Connection

The usage of solder bumps within the Controlled-Collapse Chip Connection (C4) process is actually the dominant interconnection technology at flip chip bonding (1, 3). The C4 process consists mainly of applying the UBM, forming of the solder bumps, fluxing and

Ex. 1016, Page 749

placing the chip on the substrate and final assembly reflow. An actual pitch for flip-chip chip to package mounting is 130 µm and thereby the diameter of the bump is at 80 - 90µm. To allow a further shrinkage of the whole package the pitch and the diameter of the solder bump has to shrink (1). However this leads to issues at the standard C4 process.

Non-connected joints: Because of the shape of the UBM, the wetting and surface tension of the solder the "controlled collapse" of the C4 occurs. Therefore the stand off height of the chip from the substrate after final reflow is 10 – 20% smaller than the original solder bump height. Solder bumps show a certain height variation before assembly reflow and the "controlled collapse" guarantees the contact of all. The standard deviation of the height doesn't shrink in the same way with the bump size as the collapse height. When the two values are of the same size, the assembly yield might drop, because the C4 process doesn't guarantee the contact of the interconnection pads with the solder any more, which is the case for solder bumps below 50µm diameter (solder ball diameter and standard deviation in the literature: 45 ± 4 µm, 28 ± 3µm, 25 ± 4µm) (4 - 6).

The issue is enforced by the usage of thin chips because the mechanical strength of the silicon is lower and they can show warp because of the functional layers, which are deposited usually at elevated temperature and have different CTE compared to silicon (7). Additionally the temperature cycling while assembly reflow can induce further dynamic warp which again leads to open joints because of the small compliant capacity of the small solder balls (2).

The application of force while the assembly reflow can reliably ensure the contact of all solder balls (figure 1). The desired stand off height can be ensured and squeeze out of solder avoided by movement limiting features like high melting point balls as already used in some the solder bumps.

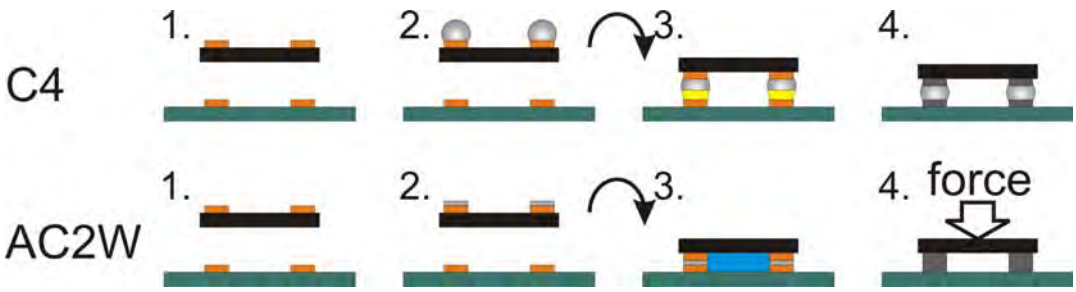


Figure 1. C4 and AC2W (for Cu-SnAg-Cu eutectic bond) process flow comparison:
C4: 1. applying of under bump metallization, 2. applying of solder with ball forming reflow, 3. fluxing and placing on substrate, 4. assembly reflow;
AC2W: 1. applying of under bump metallization, 2. applying of solder without ball forming reflow, 3. aligning and tacking on substrate by temporary adhesive, 4. assembly reflow with force

High portion of intermetallic compound: The UBM separates the copper from the solder only for limited time and an intermetallic compound (IMC) forms spontaneously between the UBM and the solder. IMC, which consist of metal complexes like Cu_6Sn_5 or Cu_3Sn, forms by diffusion processes and grows further even at room temperature. Solder joints are therefore in a thermodynamical non-equilibrium. The ongoing growth of IMC causes failures and a huge amount of research was done in this area especially on brittleness and electromigration (8, 9). In the final solder joints there are three structural

Ex. 1016, Page 750

layers, which are the bottom IMC, the solder and the top IMC. Since the growth kinetics doesn't depend directly on the bump size the portion of the IMC layers rises with shrinking bump dimension and for bumps smaller 50µm it can reach 20 % over the lifetime.

Therefore with shrinkage of solder bumps the C4 process migrates to a CuSn eutectic bond process when all tin is consumed and converted to an IMC. Consequently the joint consist of only one structure, the IMC and when sufficiently cured it is in a thermodynamical stable state, the Cu_3Sn IMC. It was shown that this kind of connection has several electrical and mechanical advantages compared to non-completely converted solder joints, as low resistivity, no electromigration, high shear strength and high mechanical stability (13, 14). Publications show the feasibility of this kind of chip-to-substrate and chip-to-chip connection, which fulfills required reliability tests (10 - 13). Compared to C4 the amount of solder is strongly reduced and adjusted to the copper volume (copper 5µm, solder SnAg 1.5µm) (11). A first reflow for bump formation is not necessary and the missing compliant capacity of the solder joints is overcome by application of force while final assembly reflow which not only has the function to bring the connection pads in contact but is also necessary to form a good eutectic bond.

As shown above for small C4 interconnects application of force while final reflow can avoid low assembly yield by open joints. At other chip interconnections methods, where also solder is used as connection forming material, like copper or gold pillars, the same effects occur at low solder volume and therefore adjusting the process in the mentioned manner can also improve the assembly yield (15).

In the following the Advanced Chip to Wafer (AC2W) process flow will be shown (18, 19). It enables to do the assembly reflow step with application of force while the alignment of the chips is preserved. The AC2W bonding technology is suited for all bond processes where force is applied while the process, like direct metallic, thermocompression and eutectic bonding. These bonding methods are not only suitable for chip to substrate connections but also for chip stacking and 3D integration (11). All bonding methods which are known from wafer to wafer bonding can also be performed on chip to wafer level by the AC2W bonding process.

AC2W bonding technology

With the Advanced Chip-to-Wafer bonding technology single components (electronic dies, odd shaped components, etc.) can be mounted on substrates. This process was developed especially for mounting electronic chips on semiconductor wafers and processes, where force is necessary to guarantee a good connection and bond formation. As showed in figure 2.a. the main idea behind the AC2W bonding process is to split up the flip chip process into two steps, the aligning step and the permanent bonding step, in a comparable way as it is done very successfully at wafer to wafer bonding. This leads to a high overall throughput for the AC2W bonding process flow.

Alignment and Tacking of Chips on the Wafer: In the first step of the AC2W process the chips are aligned and tacked to the substrate. This can be done on a standard flip chip bonder (figure 2 b.). The process fixes the alignment of the chips for the transport of the chip-substrate stack from the flip chip bonder to the chip-to-wafer bonder where all chips are bonded in parallel at once permanently on the wafer. Actually there are four different possibilities of tacking the chips after the alignment: flux dipping, high temperature evaporating adhesive, no-flow underfill and ultrasonic tacking. At flux dipping a small

Ex. 1016, Page 751

amount of flux is applied on the solder bumps by dipping it into a bath of flux. This process is similar to the standard process used for C4. For high temperature evaporating adhesive different kind of materials can be used, as bibencyl or polyethylenglycol. The materials act as an adhesive at low temperature but evaporate without residues at bonding temperature. No flow underfills, also well known from die attach processes, can also be used as temporary adhesive and in contrast to the evaporating adhesives stay there for mechanical stabilization. If any organics should be avoided chips can be temporarily mounted by ultrasonic tacking where the metal contacts themselves form a low adhesion by introduced ultrasound. None of these tacking processes are more time consuming then the standard flux dipping and therefore the flip chip bonder can tack the dies with high throughput.

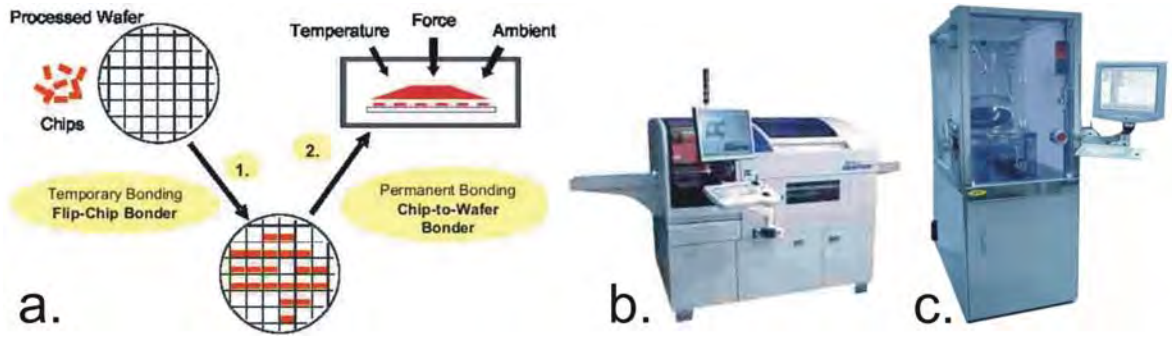


Figure 2. a. Principle of Advanced Chip to Wafer process flow: 1. aligning and tacking the chips on the wafer, 2. forming the permanent bond; b. Example of high speed flip chip bonder for AC2W bonding step 1 (16); c. Chip-to-Wafer bonder for AC2W bonding step 2 (17);

Forming the permanent bond: To create the permanent bond the fully populated wafer is transported from the flip chip bonder into the chip-to-wafer bonder (figure 2.c.). Then the final bond is made for all chips in parallel in one single process cycle. While bonding various process parameters can be varied, e.g. force on the dies, center of force, temperature, vacuum, process gasses (figure 3.a.). Certain features allow the parallel bonding of multiple chips at the same time on a substrate: the compliant layer and the center of force movement. The chips show a height variation in the micrometer range, because they are usually not from the same source wafer. To ensure the same force on each chip while bonding, the bonder has an integrated compliant layer that compensates the thickness variation. These compliant layers consist of soft material and types are available which withstand temperatures up to 550°C. When the chips are tested before tacking on the wafer and only “Known Good Dies” (KGD) are used, usually the center of gravity of the chip population is not in middle of the wafer (figure 3.b.). When force is applied on the chips, the center of force has to be shifted to the center of gravity of the chips to preserve the alignment and make a conformal bonding.

Flip chip bonder only - AC2W bonding process throughput comparison

As shown above for interconnect technology below 50µm diameter it is for several reasons necessary to apply force while bonding chips on substrates. The process can be performed either by a flip chip bonder only or by the AC2W process. At the flip chip bonder only process the steps are done in serial, so each chip is taken from the source,

Ex. 1016, Page 752

aligned and immediately permanent bonded. At the AC2W process the aligning and bonding process step are done in parallel on different tools. As an example a Cu-SnAg-Cu bonding process with 1000 chips per wafer, permanent bonding at 250°C and alignment accuracy of 7µm (3σ) is assumed. Actual numbers for the time needed for the individual steps are 0.5s for alignment of an individual chip on a substrate and 20s for the forming of the permanent bond, which represent actual state of the art values (11). However these values can vary depended on exact process parameters. The time for the permanent bond, which is compared to the alignment time 40 times bigger, is necessary to raise the temperature to the bonding temperature and to allow the metals to form the IMC. Diffusion processes are involved that have certain time constants. The throughput will be 0.175 wafers per hour (= 3600s / ((20s + 0.5s) * 1000)) when the whole process is done on a flip chip bonder.

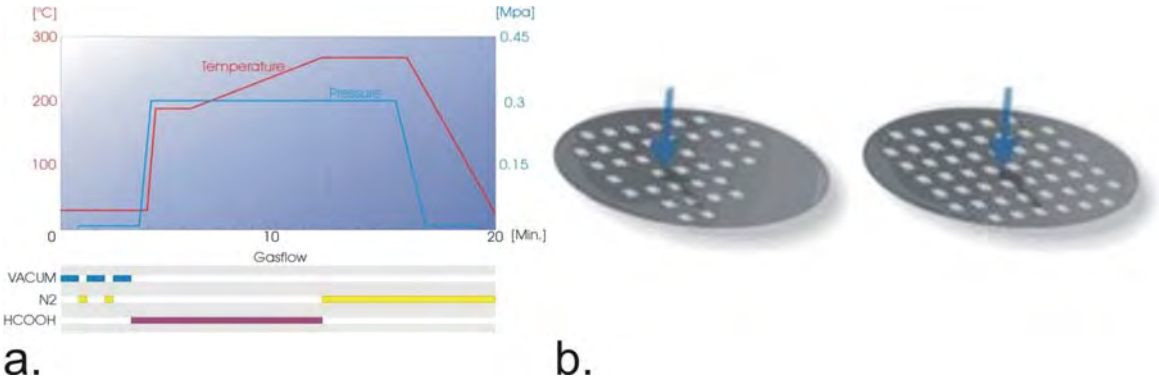


Figure 3. a. Example process parameters for eutectic bond (19); b. Shifted contact force due to population of only “Known Good Dies” (KGD);

At the AC2W process the alignment and tacking of all 1000 chips is done within 500s (= 0.5s * 1000) on the flip chip bonder and the permanent bonding process needs 20min (= 1200s) at the EVG540C2W chip to wafer bonder, which again can vary and is a process and tool dependent value for the complete permanent bonding cycle (e.g. figure 3.a.). Since two tools are involved, the two steps can be done in parallel. So the flip chip bonder can tack chips on the next wafer while the permanent bond process is in progress. Therefore the overall throughput is in this case determined by the permanent bond process with 3 wafers per hour (= 3600s / 1200s), which shows that a much higher throughput can be achieved by the AC2W process.

Conclusion

At further shrinkage of C4 solder joints assembly yield drops because of open joints due to solder bump or layer thickness variations and chip warpage. This is avoided by application of force while assembly reflow. Additionally shrinkage of the C4 is equal to a migration to eutectic bonding interconnection technology where force while bonding is necessary to form a proper bond.

When the bond process with the application of force at assembly reflow is done on a flip chip bonder only the throughput is very low because of heating cycles of the tool and time consuming diffusion processes in the material. At the AC2W bonding process the two flip chip steps, aligning and the permanent bonding, are separated. The big throughput advantage of this concept was shown.

References

1. ITRS, "Assembly and Packaging 2007", *http://www.itrs.net* (2007)
2. ITRS, "The next Step in Assembly and Packaging: System Level Integration in the package (SiP)", *http://www.itrs.net* (2007)
3. L. F. Miller, *IBM J. Res. & Dev.*, **13**, p. 239 (1969)
4. K. Tatsumi, K. Shimokawa, E. Hashino, Y. Ohzeki, T. Nakamori and M. Tanaka, *The International Journal of Microcircuits and Electronics Packaging*, **22**, No. 2 (1999)
5. D. Manessis, R. Aschenbrenner, A. Ostermann and H. Reichl, *33rd International Electronics Manufacturing Technology Conference*, p. 1 (2008)
6. J. U. Knickerbocker, P. S. Andry, L. P. Buchwalter, A. Deutsch, R.R. Horton, K. A. Jenkins, Y. H. Kwark, G. McVicker, C. S. Patel, R. J. Polastre, C. Schuster, A. Sharma, S. M. Sri-Jayantha, C. W. Surovic, C. K. Tsang, B. C. Webb, S. L. Wright, S. R. McKnight, E. J. Sprogis, B. Dang, *IBM J. Res. & Dev.*, **49**, No. 4/5, p. 725 (2005)
7. B. Dang, S. L. Wright, P. S. Andry, E. J. Sprogis, C. K. Tsang, M. J. Interrante, B. C. Webb, R. J. Polastre, R. R. Horton, C. S. Patel, A. Sharma, J. Zheng, K. Sakuma and J. U. Knickerbocker, *IBM J. Res. & Dev.*, **52**, No. 6, p. 599 (2008)
8. P.-E. Tegehall, *IVF Project Report 06/07, http://www.ivf.se* (2006)
9. K. N. Tu, *Journal of Applied Physics*, **94**, No 9, p. 5451 (2003)
10. K. Sakuma, P. S. Andry, C. K. Tsang, S. L. Wright, B. Dang, C. S. Patel, B. C. Webb, J. Maria, E. J. Sprogis, S. K. Kang, R. J. Polastre, R. R. Horton and J. U. Knickerbocker, *IBM J. Res. & Dev.*, **52**, No. 6, p. 611 (2008)
11. M. Umemoto, K. Tanida, Y. Nemoto, M. Hoshino, K. Kojima, Y. Shirai and K. Takahashi, *Electronic Components and Technology Conference, Proceedings 54th*, **1**, p. 616 (2004)
12. S. Pozder, A. Jain, R. Chatterjee, Z. Huang, R. E. Jones, E. Acosta, B. Marlin, G. Hillmann, M. Sobczak, G. Kreindl, S. Kanagavel, H. Kostner and S. Pargfrieder, *Interconnect Technology Conference, IITC International*, Burlingame, CA, USA, p. 46 (2008)
13. H. Huebner, O. Ehrmann, M. Eigner, W. Gruber, A. Klumpp, R. Merkel, P. Ramm, M. Roth, J. Weber and R. Wieland, *Proc. Advanced Metallization Conference, Materials Research Society*, p. 53, San Diego (2002)
14. R. Labie, W. Ruthooren, K. Baert, E. Beyne and B. Swinnen, *Interconnect Technology Conference, IITC*, p. 19 (2008)
15. M. Hutter, M. Klein, G. Engelmann and H. Oppermann, „Investigation of Solder Bumps, Flip Chip Assembly, Reliability and Passive Alignment Using Au/Sn", *Anniversary Edition, 60th Birthday of Herbert Reichl 12th of January*, Berlin, p. 1 (2005)
16. Homepage of Datacon, *http://www .datacon.at*
17. Homepage of EVGroup, *http://www .evgroup.com*
18. C. Scheiring, H. Kostner, P. Lindner and S. Pargfrieder, „Advanced-Chip-to-Wafer Technology: Enabling Technology for Volume Production of 3D system integration on wafer level", *IMAPS 2004, Long Beach*, 14 – 18 November (2004)
19. S. Pargfrieder, T. Matthias, C. Schaefer, M. Wimplinger, C. Scheiring, H. Kostner, „3D Packaging via Advanced-Chip-to-Wafer (AC2W) bonding enables Hybrid System-in-Package (SiP) Integration", IWLP – Wafer-Level Packaging Congress, November 2-4, San Jose, California (2005)

ECS Transactions, 18 (1) 727-731 (2009)
10.1149/1.3096528 © The Electrochemical Society

Material Solutions for Advanced Thin Packages

Hiroyuki Tanaka

Sumitomo Bakelite Co. Ltd.
Information and Telecommunication Material Laboratories
20-7, Kiyohara Industrial Park, Utsunomiya, Tochigi, 321-3231, Japan

This paper describes thermo-mechanical reliability parameters of thin flip-chip ball-grid array packages—warpage at room and reflow temperatures and solder-bump-fatigue during temperature-cycling tests—as calculated using finite element method. Viscoelasticity of the underfill is considered to improve the accuracy of calculations. The parameters obtained considering the viscoelasticity of the underfill are compared with those obtained considering elasticity alone; they are also compared with parameters obtained from experiments. Optimal properties of the underfill are proposed for thinner reliable flip-chip packages. Results of this study also demonstrate that substrates with a smaller coefficient of thermal expansion and with higher stiffness are necessary for thinner reliable flip-chip packages.

Introduction

A flip-chip (FC) package with a lead-free solder bump is being introduced to meet demand for semiconductor packages with high-speed and multiple functions. Thinner and lower-profiled packages are necessary in FC packages for hand-held products such as mobile telephones and digital cameras. Furthermore, many plastic materials are applied as packaging materials in FC packages. In this study, thermo-mechanical characteristics of thin FC packages related to its reliability are analyzed against the physical properties of packaging materials using finite element method (FEM). Viscoelasticity of the underfill is considered for detailed calculations in the analyses.

Analysis

Thermo-mechanical characteristics of the FC package related to its reliability

For FEM analyses, calculations of many kinds are necessary to clarify the thermo-mechanical reliability of FC packages. Calculations were done for an FC ball grid array (FC-BGA) package in these analyses. Warpage at both room and reflow temperatures and solder bump fatigue during temperature cycling processes were calculated against properties of packaging materials for the FC-BGA package. The structure of the FC-BGA package used for these analyses is portrayed in Fig. 1 along with the solder bump joint structure. Underfill is usually applied as an encapsulant for bump joints. The Si chip dimensions are 15 mm × 15 mm × 725 μm. The package substrate comprises a substrate core with six build-up layers on it. The core thickness is 400 μm. The substrate size is 50 mm × 50 mm. The fillet length of the underfill is 1.7 mm. Software (Marc 2007 R1; MSC Software Corp.) was used for FEM analyses. Temperature dependences of physical properties for all packaging materials were considered in FEM analyses, along with the viscoelastic properties for the underfill and elastic-plastic-creep properties for the solder bump.

Ex. 1016, Page 755

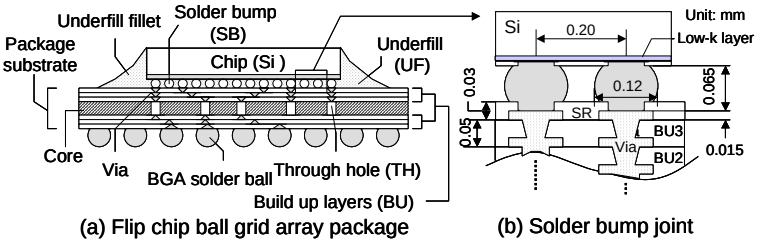


Fig. 1 Structures of the analyzed flip chip package and solder bump joint

UF properties

Temperature dependences of the moduli for the underfills used in the FEM calculation are presented respectively in Fig. 2(a) and Fig. 2(b): Underfill series A – A, AE1H, AE1L, AE2H, AE2L; Underfill series B – B, BE1H, BE1L, BE2H, BE2L. Temperature dependence of CTE for the underfills of series A and B are presented in Fig. 2(c). The T_g values are 45°C for series A and 75°C for series B.

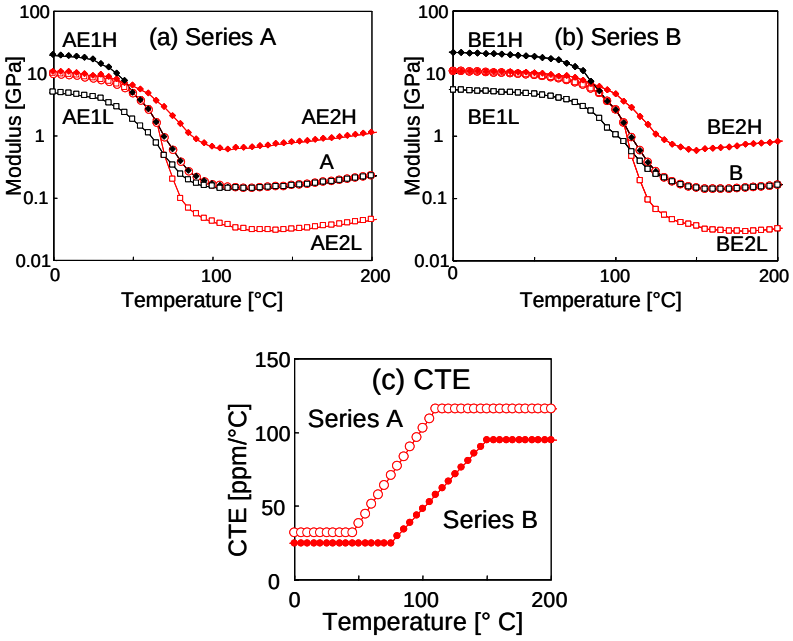


Fig. 2 Physical properties of underfills used in analyses:
(a) Series A modulus, (b) Series B modulus, and (c) CTE

Viscoelastic model used for calculations

Relaxation moduli and their temperature dependences for the underfills were determined from temperature dependences of the storage moduli shown in Figs. 2(a) and 2(b) using a method described in a prior study (1). As an example, a comparison of global warpage values calculated using the elastic analysis to viscoelastic ones for a FC-BGA package is depicted in Fig. 3. The warpage values calculated using the two analysis methods differed greatly from each other at room temperature, confirming that the warpage values calculated using results of viscoelastic analysis agreed well with those for actual FC-BGA packages in the experiment.

Ex. 1016, Page 756

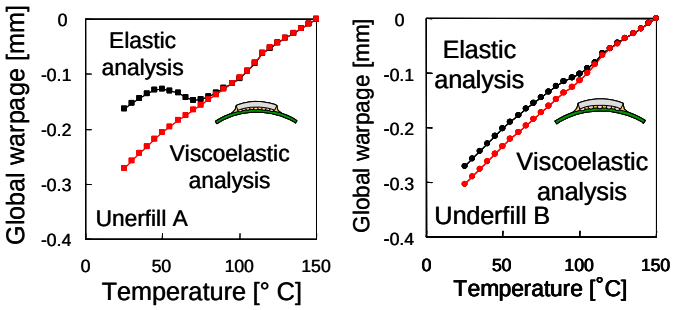


Fig. 3 Comparison of global warpage obtained using different analysis methods

Warpage analysis

Global warpage values calculated at 25°C for underfills are portrayed in Fig. 4(a). No elasticity dependence of warpage is shown, except for T_g dependence. Detailed analyses revealed that stress relaxation in the fillet of the underfill strongly affected global warpage for FC packages. Global warpage values calculated at a temperature of 260°C, which corresponds to the maximum temperature during the assembly process, are presented in Fig. 4(b). The global warpage values at 260°C depend not only on the T_g value of the underfill, but also on the modulus. The global warpage values of underfill A are larger than those of the AE2H or AE2L underfill, which has larger or smaller modulus characteristics at temperatures above *Tg* than that of underfill A. It is also similar for underfill B. Warpage characteristics at 260°C are complex against the underfill properties.

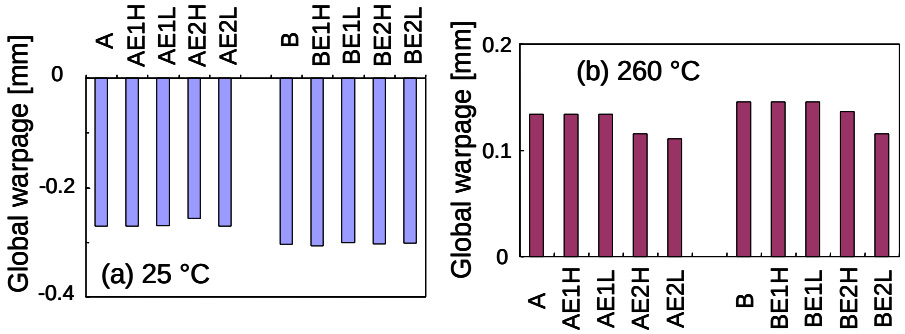


Fig. 4 Global warpage values calculated at (a) 25°C and (b) 260°C

Analysis of solder bump fatigue during temperature cycling

Figures 5(a) and 5(b) show the maximum inelastic strain of the solder bump calculated against time during a temperature cycle test using temperatures from -55°C to 125°C for the FC-BGA package. As these figures portray, inelastic strains obtained from viscoelastic analyses for underfills A and B are 3–5 times as large as those by the elastic analysis. Figure 6 depicts the inelastic strain range calculated for the model underfills. The inelastic strain range is defined as the increase in inelastic strain in a period of the temperature cycle. The inelastic strain range is smaller for underfills having larger *Tg* values and having larger modulus values at temperatures higher than *Tg*. Results show that underfills having smaller *Tg* and higher modulus at above *Tg* are appropriate for use in thin FC packages.

Ex. 1016, Page 757

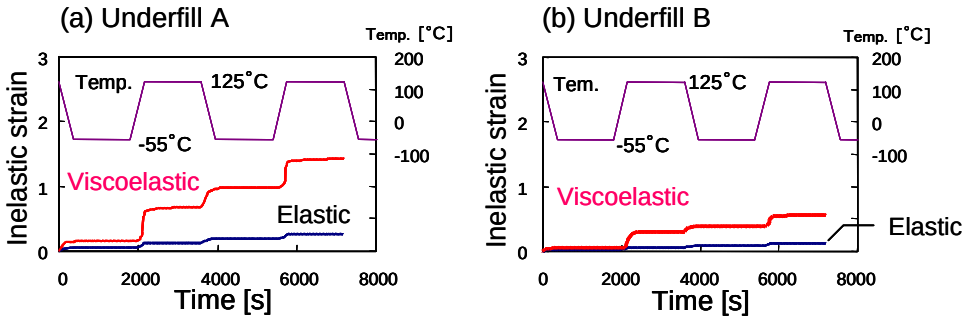


Fig. 5 Inelastic strain of the solder bump in a temperature cycle of -55°C to 125°C for underfills A and B

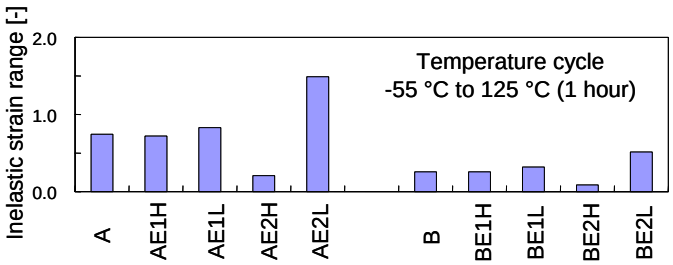


Fig. 6 Inelastic strain ranges of the solder bump

Influence of package substrate properties

Reliability of the FC package was also affected by the package substrate properties. Figure 7 portrays the warpage calculated for an FC-BGA having a thin conventional substrate with a 0.1-mm-thick core, as compared to that with a 0.4-mm-thick core. Warpage of the FC-BGA package made of conventional substrate materials increases as the substrate thickness decreases. The warpage characteristics for a FC-BGA with a thin substrate having smaller CTE and higher stiffness core are shown in the figure. The warpage increases very little as the core thickness decreases for packages with smaller CTE and higher stiffness core. Figure 8 depicts solder bump fatigue characteristics for the package with thin core with smaller CTE and higher stiffness. The inelastic strain range does not increase as the core thickness decreases for the core with smaller CTE and higher stiffness.

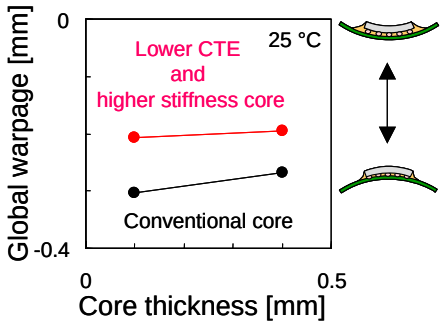


Fig. 7 Calculated warpage for FC-BGA packages with a thin substrate core

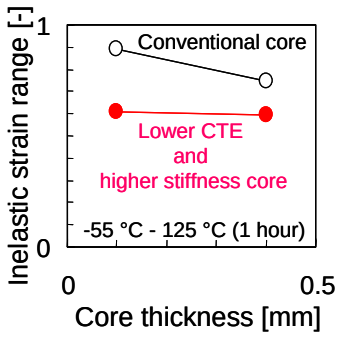


Fig. 8 Calculated inelastic strain ranges of a solder bump for FC-BGA packages with a thin substrate core

Conclusion

Effects of physical properties of packaging materials on the thermo-mechanical reliability of thin flip chip packages equipped with a lead-free solder bump were analyzed parametrically using FEM considering the viscoelasticity of underfills. Results show that it is necessary to consider viscoelasticity of the underfill for precise analyses of package reliability. A substrate with smaller CTE and higher stiffness used for flip chip packages improved reliability without decreasing T_g of the underfill.

References

1 K. Miyake, *Thermo-viscoelastic analysis for warpage of ball grid array packages taking into consideration of chemical shrinkage of molding compound*, JJIEP, Vol. 7, No. 1, pp. 54–61 (2004).

ECS Transactions, 18 (1) 733-737 (2009)
10.1149/1.3096529 © The Electrochemical Society

Development of Epoxy-Silicone Hybrid Polymer and Application for Semiconductor Assembly

Kazutoshi Tomiyoshi, Tatsuya Kanamaru, Kenji.Fukui
Silicone Electronics Materials Research Center, 3rd Section
Shin-Etsu Chemical Co., Ltd.
Adress:1-10 Hitomi, Annaka, Gunma, 379-0224, Japan
k_tomiyoshi@shinetsu.jp

Abstract

Semiconductor package style is moving to 3D packaging and thickness become thinner. As a result request to assembly materials is to lower the internal stress for reducing mechanical damage to silicone die. Epoxy-Silicone hybrid technology has been developed for reducing stress of epoxy materials. It has been successfully applied to epoxy molding compound, underfill materials and B-stage die bonding materials as well.

Introduction

Epoxy composites have a long history of use as protective materials or adhesive materials for semiconductor devices because of its good performances like higher adhesion to substrates, higher mechanical strength and faster curability. On the other hand one of disadvantage of Epoxy resin is stiff property of cured matrix. So already some lower stress Epoxy composites modified with silicone polymer has been developed and has been applied to semiconductor field.

However in recently further stress sensitive semiconductor package has been developed and its quantity of production are increasing. For example 3D packaging, wafer level package and large module for automotive [1) 2) 3)] (Figure 1). Therefore much low stress encapsulation materials and die attach adhesive are looking for.

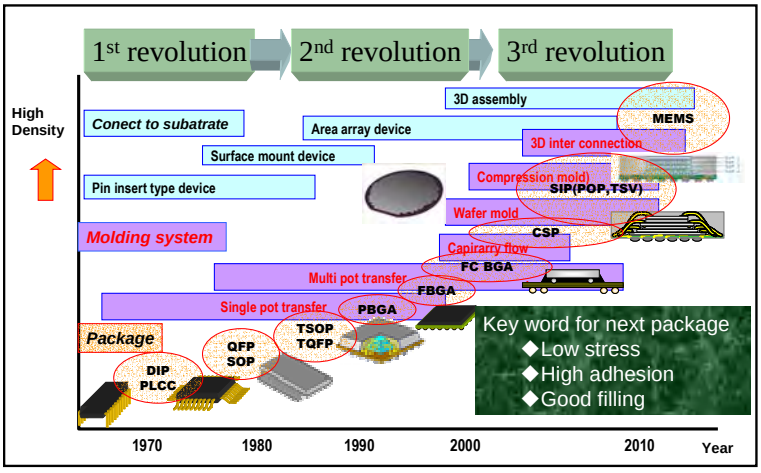


Figure 1. Package & Assembly Development Trend

Shin-Etsu has a technology of lower stress of epoxy resin by hybridization with

silicone polymer. So far unique Epoxy-Silicone hybrid polymer has been developed and it has been successfully applied to encapsulation materials and applied to B-stage die bonding materials as well.

In this paper, hybridization technology and its latest application are introduced.

1. Hybridization technology

Epoxy polymer, silicone polymer and Polyimide polymer is using in semiconductor assembly field. Those polymers have each advantage properties. Furthermore to get much more performance hybrid polymer can fulfill among this area (Figure 2).

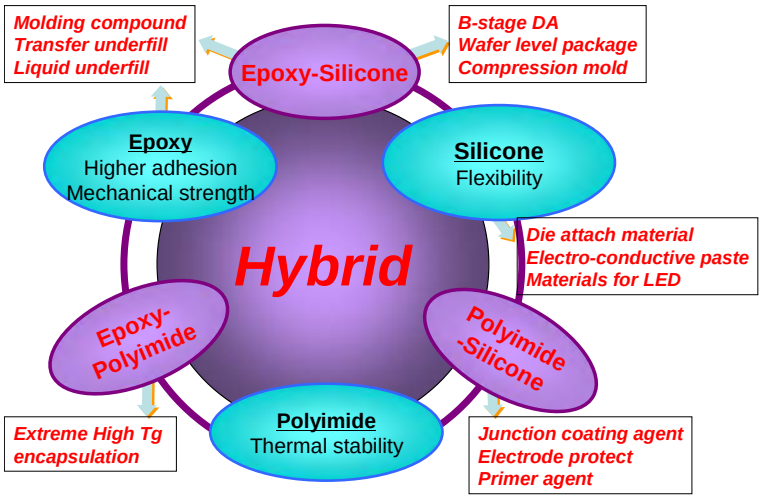


Figure 2.Hybrid polymers and total solution for semi-conductor assembly

Epoxy-Silicone hybrid polymer is synthesized with organic part and silicone polymer. Figure 3 shows example of Epoxy-Silicone hybrid model & cured matrix model.

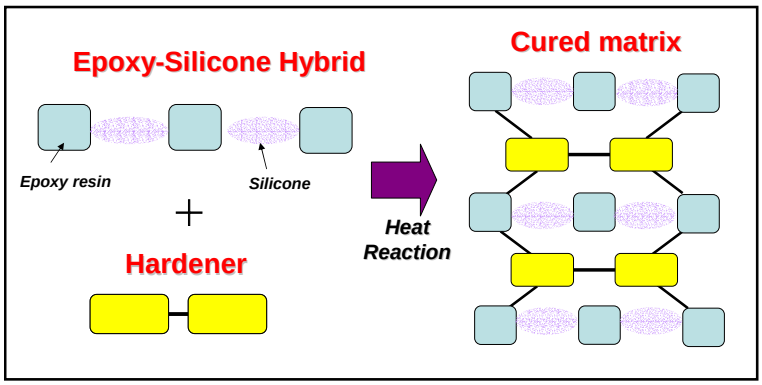


Figure 3. Epoxy-Silicone Hybrid model & Cured Matrix Model

Shin-Estu can make several type of Epoxy-Silicone hybrid structure. And can change modulus value & Tg value of cured matrix. Figure 4-1 shows hybrid structure type vs modulus and Tg. Figure 4-2 shows hybrid polymer content in cured matrix vs modulus & Tg. Type A-D show lower modulus than without hybrid polymer while keeping high Tg value.

Ex. 1016, Page 762

Feature of Type E is very low modulus and lower Tg. When additive content is increasing the modulus become lower than 1 GPa even if silica content is 40wt%.

Hybrid A-D is called “HIGH Tg & LOW MODULUS” type. Hybrid E is called “LOW Tg & SUPER LOW MODULUS” type.

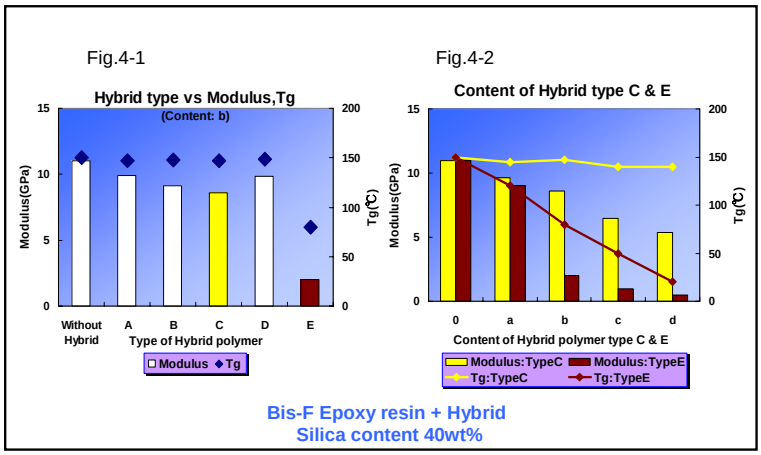


Figure 4. Hybrid polymer can control modulus & Tg of cured matrix

2. Application for semiconductor assembly

Figure 5 shows example of application for each hybrid type. Hybrid polymer can be applied to many assembly materials.

Hear B-Stage die attach adhesive and 2nd Underfill material is introduced.

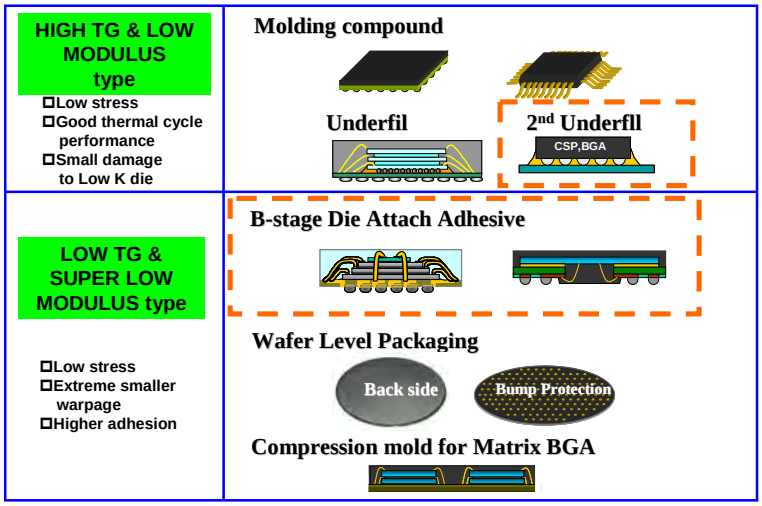


Figure 5. Application of Hybrid polymer

2-1. B-Stage Die attach adhesive

B-stage adhesive is useful for BOC-BGA, FBGA, stacked BGA as die attach and interposer of stacked die. SFX-513 series are developed as B-stage adhesive which is applied “LOW Tg and SUPER LOW MODULUS” type Epoxy-Silicone Hybrid. When mold FBGA with a molding compound, package warpage become very small. And this package can pass MSL level 1(Figure 6).

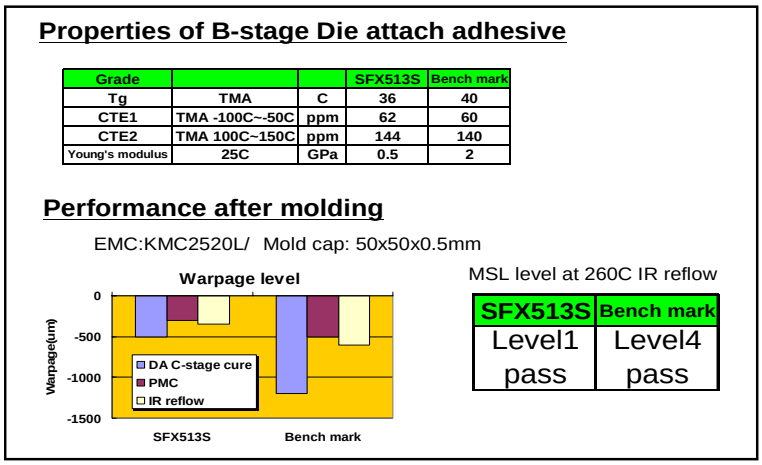


Figure 6.B-stage die attach adhesive

2-2. 2nd Underfill material

2nd Underfill is usefull for device protection. Then anti crack performance is very important property for 2nd underfill. SMC386 series is applied of “HIGH TG & LOW MODULUS” Epoxy-Silicone Hybrid. This material can cure at low temperature and has good drop test performance. And this material has good reworkability (Figure 7).

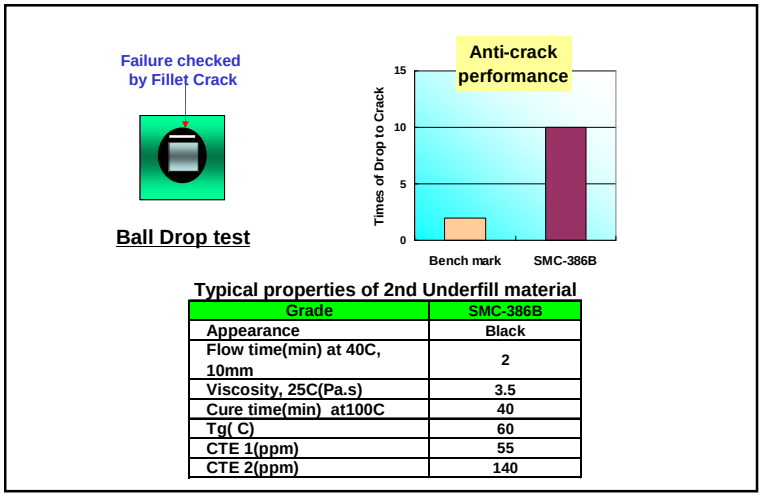


Figure 7.2nd Underfill material

Conclusion

1. Epoxy-Silicone Hybrid polymer has been developed. It was applied many semi-conductor assembly materials and can be get good low stress performance.
2. Recently B-stage die attach adhesive and 2nd underfill materials was developed and commercialized. Those materials are very useful to get high performance devices.
3. SFX-513S is B-stage die attach adhesive. The feature of this material is super low modulus and has high reliability.
4. SMC-386B is 2nd Underfill material which can cure at low temperature. And has good anti-crack performance and good reworkability.

Reference

1. Y.Hayashi et al, *Fabrication of Three-Dimensional IC using CUBIC Technology, Digest of 1990 Symposium in VLSI Technology, p95(1990).*
2. K.Hirano et al, *A New Three Dimensional Multiport Memory for Shared Memory in High Performance Parallel Processor System, Extended Abstracts of the 1996 International Conference on Solid State Device and Materials, p824(1996)*
3. K.Kusamitsu et al, *3 Dimensional Memory Module Assembly Technology, The Third VLSI Packaging Workshop of Japan,p133 (1996.12)*

ECS Transactions, 18 (1) 739-743 (2009)
10.1149/1.3096530 © The Electrochemical Society

UV CURABLE ADHESIVES FOR CMOS & CCD IMAGE SENSOR PACKAGE

Taro Kenmochi

Optical Semiconductor Group, Kyoritsu Chemical & Co., Ltd, Tokyo, Japan

The packages of CMOS & CCD image sensor are widely used to digital still camera, digital video camera, D-SLR (single-lens reflex), surveillance, automotive and others. On those applications, the packages are composed of plastics or ceramics with using UV Sealant[1-2]. UV curable sealants have contributed to increase reliability and productivity of the assembly line of CCD or CMOS packages. UV sealants necessary have specific characteristics which are air-tight, visco-elasticity and others. Kyoritsu Chemical, we offer those technical data and explanations why those data are essential. We demonstrated necessity of required properties which affected the technical issues to produce packages and proposed causes and countermeasures. We have not only supplied UV curable sealants but have offered total solutions for the production of packages.

Introduction

UV curable sealants have been used CCD, CMOS packages in market. Currently, image sensor markets require reflow process and most assembling manufactures are trying to make reliable packages against it. In this presentation, we introduce and explain 5 issues, as follows, (1) Explanation of general process, (2) Characteristics of UV curable sealant, (3) Requirements of UV curable sealants from market, (4) Kyoritsu Chemical original evaluations & test programs, (5) Troubleshooting. Those factors which affect production quality to assemble CCD, CMOS packages and proposed suitable UV curable sealants.

CCD & CMOS Image Sensor Assembling Process

The assemble lines of CCD & CMOS image sensor packages are widely varied. Depends on process, UV sealants should be chosen to meet the demands. Those are consisted of dispensing of UV curable sealants, attaching glass lid onto package substrates, UV cure and heat annealing (Figure 1). After completion of those processes, surface mounting process with Reflow is applied. Explain briefly what objection is each process.

Table 1 Typical assembling process of CCD & CMOS image sensor packages

Process	Objection	Effecting
Dispense	Dispense UV sealant ceramics or plastic package.	Productivity
Glass Lid attached	Air, Mechanical, Glass-weight or Vacuum	Release internal stress
UV Cure	UV sealant is cured by UV radiator	Productivity
Heat Anneal	Stress relief and Reaction rate is increased	Reliability

Ex. 1016, Page 767

·UV cure process
I. Fiber type (High pressure mercury lamp): Fit on cell-line and avoid high temperature in process.
II. Area type (Metal halide lamp): Fit in-line production and apply several packages in one lamp.

·Heat Anneal process
I. Hot plate: Necessary to avoid rapid increase of temperature.
II. Furnace: Temperature is very high and sometime affects to stress in package.

Those processes have each characteristic and need optimum conditions. If process line is not suitable, reliable packages cannot be obtained.



Figure 1 Schematic illustration of process line

UV Sealant Characteristics

The table 2 shows data of Kyoritsu Chemical UV sealants including viscosity, DMA, HIC, water vapor permeability, adhesives strength and others. Viscosity affects process ability in production line. Basically, 30000-80000mPa·s is very common to use. High viscosity will be effective to reduce stress after cured. On the other hand, low viscosity will increase throughput but will be to prevent delamination after JEDEC Reflow process. Therefore, DMA for visco-elasticity measurement is very essential for choosing UV sealants. Visco-elasticity contributes cured film between low temperatures from high temperature. Transition point of tanδ also helps choosing UV sealant. To check moisture content is necessary for high resolution applications. We are able to provide moisture data. Depends on application and package size, UV sealant will be chosen differently.

Table 2 Technical data of World Rock® UV sealants in Kyoritsu Chemical

	W/R No. 8723K9C	W/R No.8723K9B	W/R No. 8724L2	W/R No. 8724R3L
Main material	Epoxy	Epoxy	Epoxy	Epoxy
Viscosity (mPa· s)	36,000+/-7,000	37,000+/-7,000	49,000+/-10,000	39300~
Thixotropic	2.2	1.9	1.8	2.4
UV cure	6J	6J	6J	6J
Heat anneal	120℃ x 30min	120℃ x 30min	120℃ x 30min	120℃ x 30min
Water Vapor Permeability 85℃/85% (g/m²· 24hr)	6.93	6.49	7.14	5.3
Water Absorption 25℃ x 168hr (%)	1.2	2.1	1.2	1.0
CTE (a1) ppm < Tg	67	57	53	62
CTE (a2) ppm > Tg	190	160	150	232
DMA (Tg) ℃	137	150	143	109

There is additional information as below
DMA Visco-elasticity (tan δ): critical for JEDEC Reflow requirement.
We have other technical data as HIC data, Outgas HIC data, GC-MS, TG-DTA.
These data will highly affect the technical issues.

Ex. 1016, Page 768

Requirements of UV curable sealants for image sensor market

Current markets' demands are high resolution pixel size and high reliability package. UV sealants are also required high quality. World Rock® UV sealants provides low stress and can be applicable for JEDEC Reflow type. Low stress types of sealant affect to prevent delamination after JEDEC Reflow. The package appearance is also critical issue. In addition, package appearance is very essential and end-users are care look nice package. SMT level has been increasing currently and solder process temperature also increasing. Kyoritsu Chemical also provides low outgas and low moisture content type and prevent cloudiness after assembled process. Kyoritsu Chemical is able to provide high quality UV sealant. Almost of all market may use Kyoritsu Chemical UV sealant in image sensor market.

Kyoritsu Chemical original evaluation & test program

There are lot of problems and issues in the production line. Problem solutions are very important in the market. We have evaluation program so as to customer's satisfaction. We are able to provide important data as HIC, outgas-HIC, Reflow simulation, adhesive strength and others. These data help problem solution and find actual causes. We are able to support customers strongly, because we know each process well and proposes hypothesis.

Table 3 Kyoritsu Chemical Test Program.

Test	What it can measure
Adhesive Strength	Glass / Glass or Ceramics / Glass
DMA Measurement	Tan δ point (Tg), Elasticity
GC-MS	Finding Outgas during UV cure
HIC	Finding No purity Ion materials
Outgas HIC	Finding outgas within F^-, I^-, Cl^-, Br^-
FT-IR	Finding Reaction rate
TMA Measurement	Finding Tg and CTE
Water Vapor Permeability	Finding UV resin characteristics
Water Absorption	Finding UV resin characteristics
Condensation Test	Finding dew point and temperature
JEDEC Reflow Test	Checking Appearance of Delamination
JEDEC Reflow Test	Checking Airtight Characteristics

Currently, assembled manufacture and/or users care pad corrosion. Pad corrosion is occurred by many causes and happens simultaneously. Therefore, UV sealant is required high purity. We have developed fluorine reduction, no antimony and halogen content regulation. UV sealant is part of sub-material but we have tried to achieve ecological-friendly adhesives.

Troubleshooting

Troubleshooting and countermeasure is essential in production line because problems are occurred again and again. Therefore, We offer ideas, reasons and hypotheses. Main object is to prevent problems in production line.

Ex. 1016, Page 769

Table 4 Troubles & Countermeasures on Each Process

Process	Hypothesis Effect	Hypothesis Causes	Countermeasure
Preparing UV adhesive	Bubble	Defrost	Freeze➔Cool➔Room Temp. Using expanded polystyrene Box
	Gel / Crystal	Base material	Do not use and re-send
	Condensation	Cure condition	Freeze➔Cool➔Room Temp. Using expanded polystyrene Box
Dispense UV adhesive	Appearance NG	Bubble	Change dispense condition
	Workability NG	Break form	Bubble / Remain Resin / Nozzle treatment
		Blocked dispense	Bubble / Remain Resin / Nozzle treatment
UV cure Process (Lid-attach)	Uncured	Not enough UV power	Control UV radiation & UV lamp
		Adhesive strength	Find suitable UV condition
		Mask position	Control mask position
	Delamination	High Temp. at UV	Fit heat CF & Find suitable condition
Heat Anneal Process	Delamination Ceramics PKG surface	Stress / CTE	Re-design sealant width Evaluate Low tan δ type
	Delamination Plastic PKG surface	Stress / CTE	Re-design sealant width Evaluate Low tan δ type
	Delamination Lid glass surface	Remain Alkali Treatment	Cure obstruction by Alkali treatment Wet condition to PKG surface
	Float Lid glass	Heat Temp.	Prevent internal pressure / adjust Temp.
	Bend Lid glass	Stress / CTE	Low temp. heat anneal Prevent remove material Evaluate Low tan δ type

Delamination from package surface

Stress is recognized as cloud color in appearance. As a countermeasure UV sealant is chosen low stress type because UV sealant is soft and flexible, stress is able to be released. Especially, after JEDEC Reflow, white cloud color appears package edge.

Delamination from glass surface

It is occurred by cure inhition and low adhesive strength. It will be appeared by newton ring and/or rainbow. Countermeasure is to check residue of alkaline materials. Alkali materials may cause obstruction of curing progression, because these types of UV sealant are based on cationic polymerization.

Delamination from AR coating or IRCF coating

They may cause insufficient cure. IRCF coating block UV light of under 410nm. Cure of UV sealants is need to attach wavelength of 220-385nm. Therefore, it is necessary to consider a sealant with visible light curing. On the other hand, UV light can penetrate through AR coating. In addition, UV light with 385nm can penetrate significantly through AR coating. But it is not 100%. Therefore, UV cure should be conducted before package assembly. Both AR and IRCF coating is vapor deposition. It is also take care of glass and both coating strength.

Ex. 1016, Page 770

We have Reflow simulator (Figure 2) and can check suitability of UV sealant. This evaluation program is highly beneficial to choose material internally. This is critical issue UV sealant is applied JEDEC Reflow. If UV sealant is unable to pass evaluation, we are able to find causes easily and develop material. Therefore, we always meet customer's satisfaction by providing problem solution to customers.



Figure 2 Reflow simulator equipment in Kyoritsu Chemical

ACKNOWLEDGMENTS

We appreciated for all of customers using WORLD ROCK® and have been corporate UV curable sealants developing and keep maintain them based on market demands, requirements, and needs. Especially, I have very gratitude machineries and related person in R&D departments of Kyoritsu Chemical.

REFERENCES

1. Edward M. Petrie, Handbook of Adhesives and Sealants, McGraw-Hill Education , 2nd edition, United States (2007)
2. Jun Ohta, Smart CMOS Image Sensors and Applications, CRC, United States (2007)

ECS Transactions, 18 (1) 745-750 (2009)
10.1149/1.3096531 © The Electrochemical Society

Current Evolution of Wafer Thinning and Dicing

S. M. Sullivan[a]
[a] DISCO HI-TEC AMERICA, Santa Clara, California 94054, USA

Over the past few years the backend assembly process has become increasingly complex in order to add more value to products. Two processes that are linked to the changes in assembly are wafer thinning and dicing. Ultra thin stacked die and the implementation of TSVs are driving improvements to the traditional stress relief and fixed abrasive grinding processes. Both CMP (Chemo-Mechanical Polishing) and MCP (Mechano-Chemical Polishing) are being used. Wafer dicing has a number of products and processes behind its evolution. Thin and stacked die with DAF (Die Attach Film) require modifications to existing blade dicing processes. Low-k and copper wafers are requiring the use of lasers to remove materials from the dicing streets. A new and alternative laser dicing process called Stealth Dicing is an enabling technology for a number of products.

Market Drivers

Backend assembly still follows the front end that has not changed. What has changed is that much of the advancements in the capabilities of electronic devices are directly derived from innovations made in the assembly process over the past few years. Because we are more actively involved with what goes to market we add more value to products.

Thinning of wafers is now an enabling technology. Thinning is done for a number of reasons and in a number of different manners. Dicing is evolving in order to meet a variety of quality requirements along with an ever increasing number of different materials and layers being placed in the dicing street. Ten years ago a backgrinder with standard configuration and dicer with standard configuration could meet the requirements of the vast majority of devices assembled. Today that is not the case. As the saying goes, “the genie is out of the bottle”. We are no longer a commodity process.

The other side of this genie is that as a more active player it is necessary to choose directions to evolve and innovate requiring more informed decisions. Individual companies are thus more exposed to the whims of the global markets and not simply the “Silicon Cycle”.

Wafer Thinning

In the beginning most wafers were thinned to improve the performance of the device. Recently the vast majority of wafer thinning is done for form factor. Now instead of optimal silicon thickness it is simply thinner is better. This has brought about new challenges and the promise of a new technology, TSVs.

Ex. 1016, Page 773

Wafer thinning is four basic steps, grinding, handling, stress relief and cleaning. Grinding is the most basic step. Whether a wafer is ground to 500µm or 50µm thick the process is quite similar. Two fixed abrasive wheels grind the silicon. The first wheel grinds the majority of the material. The second wheel being a finer grit removes most of the subsurface damage created by the first course grind.

Si wafers become fragile as they are thinned. The force applied to a silicon wafer F_{max} to breaking is proportional to the square of its thickness t, as indicated in:

$$F_{max} = Ct^2 \quad (1)$$

The above formula makes the assumption that the silicon wafer has no subsurface damage. This is not the case with grinding. The second grind, a fine grind, greatly reduces the amount of subsurface damage but does not eliminate it. Subsurface damage reduces the strength of the wafers and die. This reduction in strength makes the handling a significant challenge within the grinder and for the processes following.

Stress Relief

Stress relief is a broad term used to remove the subsurface damage in order to regain the strength suggested by $F_{max} = Ct^2$. It also can be used as the step to expose the metal posts in TSV wafers. Dry Polishing or MCP (Mechano-chemical Polishing) and CMP (Chemo-mechanical Polishing) are the two most common forms of stress relief, followed by wet and dry etching. Each of the processes has advantages.

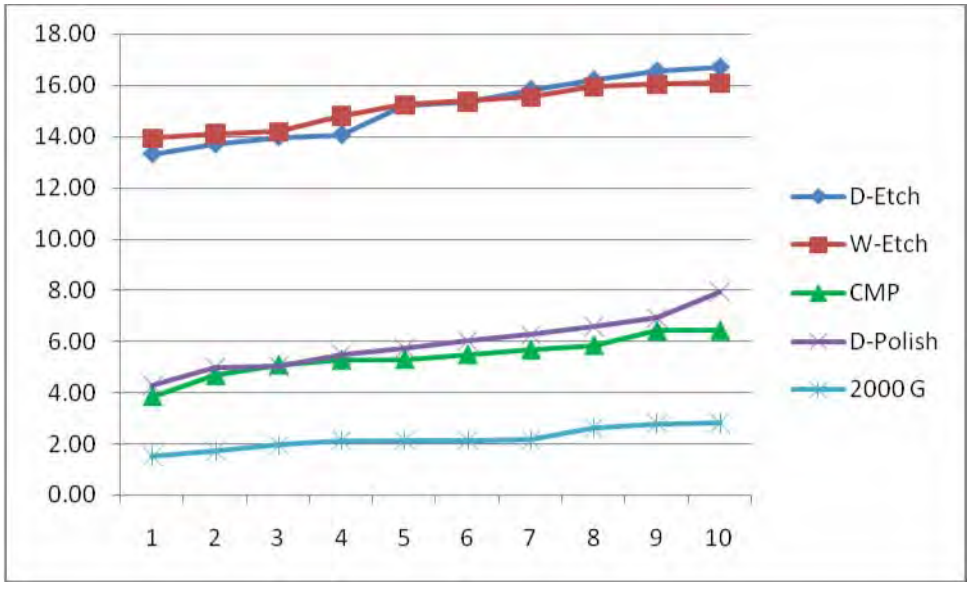


Figure 1. Relative scale of showing wafer strength Vs. stress relief method on 75µm thick product die using 3-point Bend test.

Figure 1 above shows that all stress relief methods improve die strength over the standard 2000 Grit fine grind. Wet and dry etching have the distinct advantage in die strength. Statistically there is no difference between the two etch methods for strength.

While there are technical differences between dry and wet etching often companies choose one or the other due to a familiarity with the process. Challenges associated with these two processes are well known: safety, yield, environmental and cost.

Distinctions between Dry Polishing and CMP are less well known. Again statistically there is no difference in die strength between the two polishing methods. Dry Polishing has the advantage of using no chemical slurry so all four of the above issues: safety, yield, environmental and cost are eliminated. CMP uses chemical slurry so the four issues remain, though at a much lower level than with etching. The advantage that CMP has is selectivity. When exposing metal posts in TSV wafers the slurry and pad can be chosen to selectively etch the silicon.

Though they seem quite similar Dry Polishing and CMP are quite different. CMP uses and abrasive that is harder than the silicon it is polishing. The abrasive mechanically removes the silicon and then the silicon is chemically etched removing the damage from the abrasive. Dry Polishing replaces hard abrasive which damages the wafer surface requiring the etching with a material softer than the silicon. A solid state reaction occurs at the point of contact between the substrate and polishing media. The film created by the reaction is swept away revealing a polished surface. It is important to note that because the abrasive too soft to indent the wafer no scratching or other brittle deformation occurs (2).

Challenges with Stress Relief. The surface where devices are built on the silicon wafer substrate is purified. The way that is done is by creating oxygen precipitates inside the wafer and the backside of the wafer is stressed. Over the years stress has been added to the backside of wafers in a number of ways, from glass bead blasting to laser damaging. Stress from the oxygen precipitates and backside modification create getters. These getters attract impurities at the front surface of the wafer and pull them away, creating the denuded zone(3). When wafers are ground the silicon containing the intrinsic gettering oxygen participates is removed. The thinner the silicon the less intrinsic gettering remains. This has not been a problem since the stress from the ground surface at the backside acted as an extrinsic getter and barrier. Contaminates will migrate toward or be held by areas of stress.

As wafers are thinned they lose strength. To regain some of that strength stress relief can be used. Thinning wafers removes intrinsic gettering, removing the stress from the backside of the wafer through polishing or etching removes the extrinsic gettering. Combining ultra thinning with stress relief leaves the active area of the wafer the highest stressed zone which means that contaminates will migrate to that zone.

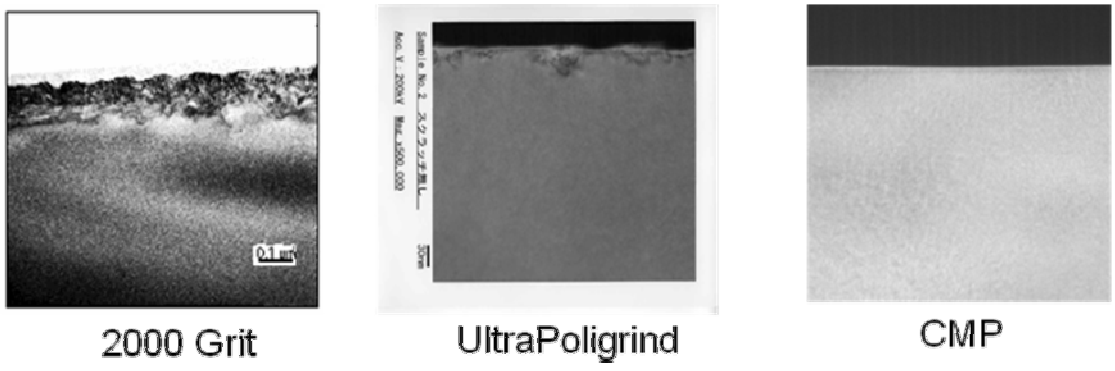


Figure 2. Amount of stress remaining at the backside of a wafer after processing

Ex. 1016, Page 775

The figures above show the stress in the silicon after grind. On the left is a standard fine grind with a 2000 grit. The depth of damage is approximately 200nm. In the middle using the ultra fine grind UltraPoligrind wheel the depth damage is greatly reduced increasing die strength while still offering a barrier to contamination. The polished surface on the right shows no depth of damage giving the highest mechanical strength but no barrier to contamination leading to poor electrical performance with some devices.

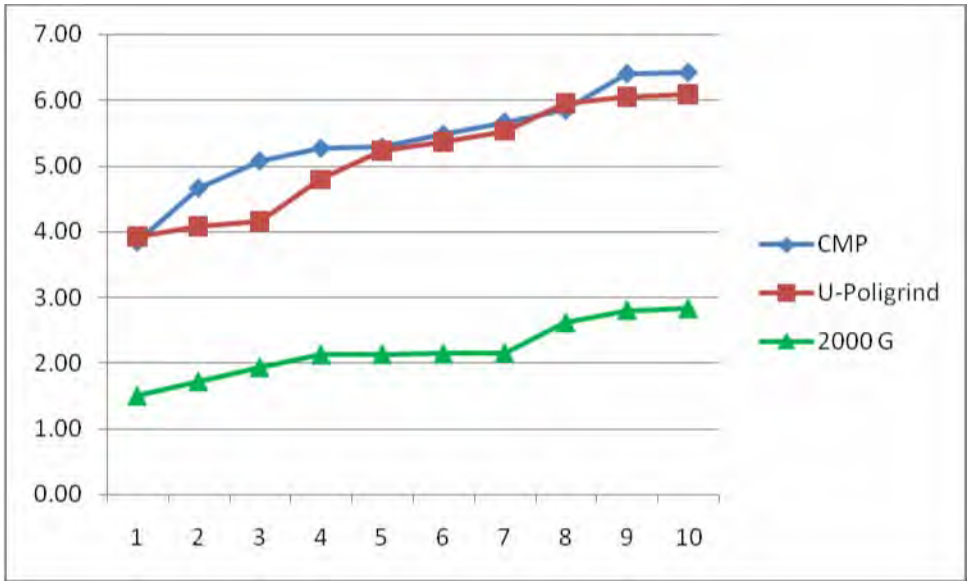


Figure 3. Relative scale of showing wafer strength Vs. process on 75µm thick product die using 3-point Bend test

TSVs

For TSVs to meet their promise wafers will have to be ultra thin. A thickness range of 5 to 20µm is commonly mentioned as targets. As of today, there is no dominate method for processing TSV wafers or die. Some TSV wafers require only standard thin wafer processes and equipment. For ultra thin via first TSV wafers there are a number of special functions being developed. As with all ultra think wafers edge trimming prior to grinding improves edge finish and yield. Bonded to a carrier wafer or on tape standard contact gauges used to measure wafer thickness can cause damage. Use of a non contact optical thickness gauge eliminates the risk of damage and accurately measures only the wafer being thinned. Third axis polishing or in line dry etching is done to recess the silicon exposing the metal posts.

Wafer Dicing

A wafer is mounted on dicing tape which is attached to a stainless steel dicing frame. The blade cuts all the way through the wafer and continues 20 to 30 µm into the tape. This process began to spread in the first half of the 1980's, and is currently the dominate method of dicing. Factors driving change in the way wafers are diced are very thin wafers, DAF (Die Attach Film), and newer materials such as low-k covering the dicing streets.

Over the past few year lasers have made a resurgence in the dicing of wafers. The applications range from scribing for a break process to removing brittle material from the saw street to full cutting.

Scribe and Break

White LEDs and some CMOS devices are being made on hard substrates such as sapphire In order to reduce the cost of manufacturing diamond scribes and diamond blades have been replaced by lasers. The quality of the scribe and therefore the break yield is equal to or better than conventional scribe. For LEDs there is typically a reduction in light output of a few percent. For the CMOS devices there is no difference in performance.

Low-k Grooving

Copper interconnect speed brought in the need for low-k dielectric materials. Lower physical strength of the low-k materials compared to conventional inter-level dielectrics create challenges during blade dicing. Associated cracking and delamination are most often addressed by slowing down the dicing feed rate and using a diamond blade that wears faster. With more advanced low-k materials this option is not available as they are too weak. Adding a laser grooving step to remove the low-k from the dicing street is the solution that is being adapted.

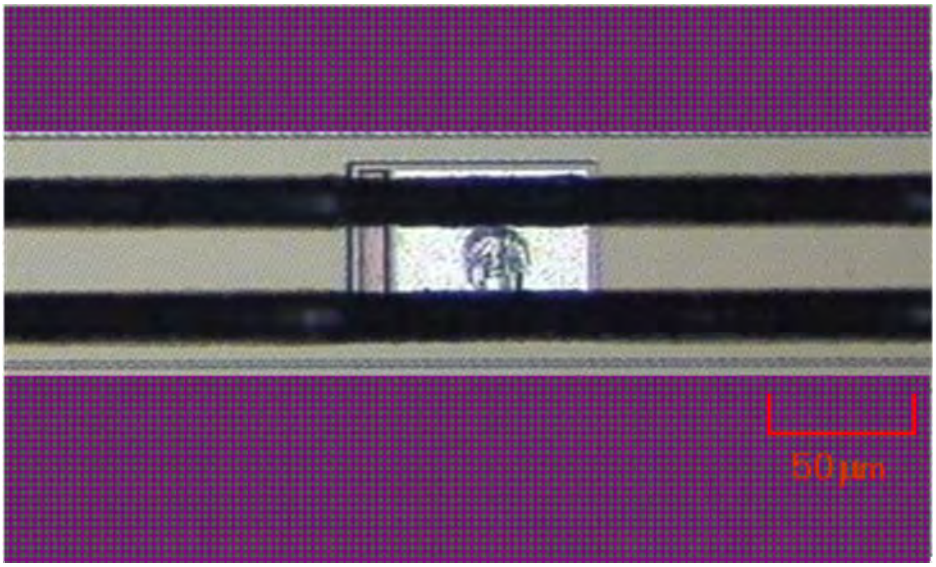


Figure 4. Two laser grooves are used to isolate blade dicing damage.

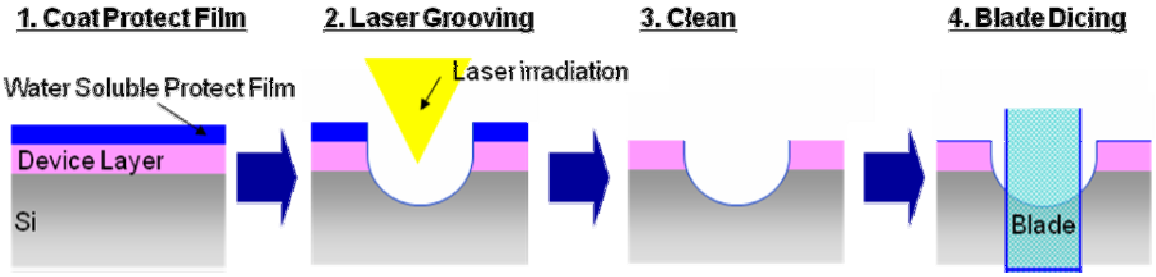


Figure 5. The laser groove is deep enough to go below the surface of the silicon prior to blade dicing.

Grooves can be two narrow or one wide one. Dicing is done between the two narrow grooves or in the center of the wide one. A wide groove removes the low-k material so that the blade never comes in contact with any. The two grooves act as breaks so that any cracking or delamination of the low-k from the blade dicing is stopped from entering the active area.

Full Cutting

Both ablation and Hamamatsu's "Stealth Dicing" can be used for full cutting of wafers. There is no clear distinction for choosing one over the other. Ablation is often applied to wafers with back side metal, DAF, non silicon wafers, or for hard materials. Stealth Dicing has the advantage of being a totally dry process and generating few to zero particles.

Stealth Dicing The Laser light with a wavelength allowing for a balance between transmission and absorption is focused below the surface of the wafer. At the focal point the energy vaporizes the material. The vaporized material expands instantaneously creating small fissures. Almost as fast the vaporized silicon solidifies into a microcrystalline structure. Shaping the laser beam in such a way that most all of the fissures are vertical creates a crack initiating layer. This process of creating the fissures and modification of the silicon creates a vertical separation between the two die. The modified layer is termed the SD layer. Wafer thickness and a number of other factors such as edge quality determine if one or multiple SD layers will be created.

Acknowledgments

I would like to thank Stewart Dunlap for his assistance and test materials.

References

1. G. Coletti, C. J. J. Tool, and L. J. Geerligs, *Poredings of the 15th Workshop on Crystalline Silicon Solar Cells & Modules: Materials and Processes* (2005).
2. N. Yasunaga, *Abrasive Technology*, p. 371 (1999).
3. R. Falster, G. Fisher, and G. Ferrero, *Apl. Physics Letters, 59 (7), 809 (1991)*
4. F. Fukuyo *et. al*., Proc. 6th Int. Symp. Laser Precision Microfabrication 2005.

ECS Transactions, 18 (1) 751-756 (2009)
10.1149/1.3096532 © The Electrochemical Society

Electroless Nickel/Gold Under-Bump-Metallization Cost Comparisons

A.J.G. Strandjord[a], A. Scheffler[a], E. Zakel[b], T. Oppert[b], and T. Teutsch[a]

[a] Pac Tech USA – Packaging Technologies, Inc., Santa Clara, CA 95050, USA
[b] Pac Tech GMBH – Packaging Technologies, Inc., 14641 Nauen, Germany

The Under-Bump-Metallurgy (UBM) is an integral part of all wafer bumping processes. This thin film layer is typically deposited by physical vapor deposition (PVD-sputtering), electroplating, or electroless plating. The electroless plating process is widely considered the lowest cost technology within the electronics packaging industry. Its implementation has been driven by the need to further reduce the costs of manufacturing. In this study, the electroless processes and their costs are analyzed using a new cost model. This cost model is based on a cost-resource structure; where the cost of materials, labor, equipment, facilities, utilities and miscellaneous costs are combined with the product mix and production schedule to calculate costs for each process step.

Electroless Nickel Process Flow

The use of electroless nickel/gold or nickel/palladium as the UBM for solder based interconnects has long been recognized as the next technology to replace both thin-film sputtered and electroplated technologies. This transition to electroless plating has been fueled by its promise of low cost processing. The basic electroless processes for plating on aluminum and copper based semiconductors have become fairly standardized over the last several years (1-2). The steps for these processes are outlined in Tables I and II.

Table I. Fab process steps.

1	Incoming Inspection
2	Back/Front Resist
3	Plasma Descum
4a	Nickel Plate
b	Palladium Plate
c	Gold Plate
d	Dry Wafers
5	Analytical Analysis
6	Final Inspection

Table II. Process steps completed within the plating line.

	Electroless Plating on Al	**Electroless Plating on Cu**
4a	Clean Organic and Si Residuals	Clean Organic and Si Residuals
4a	Etch Al Oxide and Al	Remove Cu Oxide
4a	Zincation I	Etch Cu
4a	Zinc Strip	Palladium Catalysis
4a	Zincation II	Selectively Clean Passivation
4a	Nickel Plate (2-25µm)	Nickel Plate (2-25µm)
4b	Palladium Plate (1000-2000 ?)	Palladium Plate (1000-2000 ?)
4c	Gold Plate (~200-1000?)	Gold Plate (~200-1000?)
4d	Dry	Dry

Table I lists the general fab process steps for electroless plating. Steps 1 and 6 are standard inspection steps which are common to all types of UBM deposition techniques. Incoming inspection in this study is specified within the cost model as an optical analysis at a sampling rate of 1 out of every 25 wafers, and the final inspection is selected as an automated 2D inspection at a sampling rate of 100%. Other options within the model include: 3D analysis, XRF, SEM, microscope, and profilometer.

Steps 2 and 3 are optional steps in the electroless plating flow and each has an expanded step list associated with it that is more detailed. For some wafer types, the backsides of the wafers need to be protected to eliminate any residual plating that might occur on the wafer. Examples of these include: thinned wafers, wafers with backside metal, and wafers where electric potentials may build up within the circuitry that affects the plating rate. The protection of the back side is usually accomplished by applying a blanket coating of resist, followed by an oven hard-bake. When resist protection is required on the front side, the resist steps will also include photo-patterning steps. Examples of this would be for fuse link protection, or to cover test structures in the streets. This front side protection is often already present on many wafers in the form of PI, BCB, or an Epoxy applied at the original wafer foundry or fab. All of these front side coating examples require a plasma descum step to clean the bond pads prior to electroless plating.

Ex. 1016, Page 779

Steps 4a-d are the generalized process steps for electroless plating. Step 4a includes all the pre-cleaning and nickel plating steps, while steps 4b-d are optional steps (i-Au plating, e-Pd plating, and wafer drying). Step 5 (analysis) is an important part of the electroless process, and includes both chemical analysis of the baths and test chip plating. The options for this step include: titrations, ICP, pH, and AA.

Table II expands this fab process list for step 4 into a more detailed list of operations for electroless plating. These operations are carried out as a single fab operation using an automated plating line. The plating line typically contains: load and unload stations, chemical baths which are filtered and heated, and several quick dump rinse stations. Figure 1 shows the typical bath arrangement for plating electroless nickel/palladium/gold on both aluminum and on copper based semiconductors.

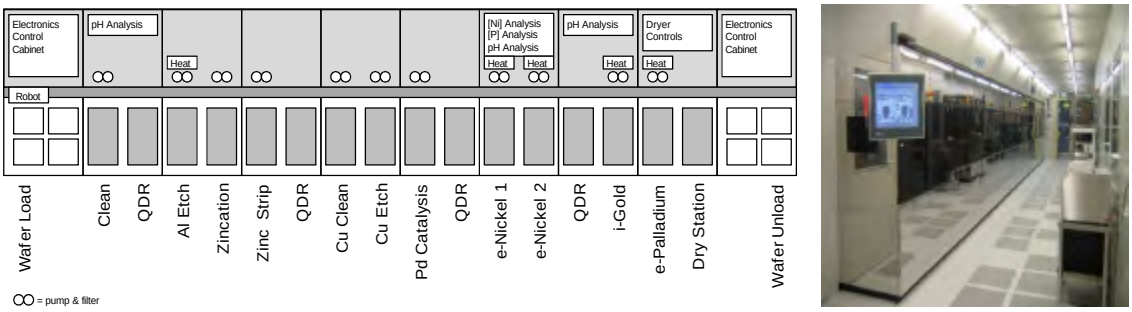


Figure 1. Schematic diagram and image of an automated electroless plating line (PacLine 3000).

The bath configuration and the degree of automation of the plating line are dependent on the wafer type and volumes of wafers being plated at the facility. The plating line shown above in Figure 1 is a plating line designed for high volume plating with a high degree of product flexibility. This line is capable of plating NiAu, NiPd, and/or NiPdAu on both Cu and Al based semiconductors. The system will automatically transport the wafers from the load station to all the selected chemical baths, QDRs, dry station, and out to the unload station using a PLC controlled robot. Wafers can range from 100 mm through 300 mm, and the line has automated chemical analysis and replenishment capabilities. Each plating-run takes approximately one hour on average to complete. The software also allows several lots to be running simultaneously to further increase plating throughput and therefore lower costs.

Cost Model

A new cost model has been developed which can be used to calculate the costs of all wafer bumping processes, including: solder paste printing, sphere drop, C4NP, sphere transfer, solder jetting, plated UBM, sputtered UBM, copper pillars, and electroless plating. The cost model is based on the Cost Resource Model developed at Sematech. The core of the cost model is a detailed process step list from which costs for each step are calculated. The steps listed in Tables I and II are examples of general step lists which are further expanded within the model to an even finer detail to accurately represent wafer movements within the fab. In many instance, several options are given for each step to allow the model to be used to calculate the cost of these different options. For example: the different options for drying the wafers include: drying within the plating line, convection oven, IPA vapor, or SRD.

The software platform for this model is Microsoft Excel and each of the boxes shown in Figure 2 represents a separate spread sheet within the model where data and variables are contained. The product mix and production schedule are used to calculate the cost for six separate “resource” categories: equipment, building, materials, utilities, labor, and miscellaneous items; which in turn are used to calculate the cost of each process step. By varying the process parameters, one can observe the effect that these changes have on the process steps and resource costs (sensitivity analysis).

Ex. 1016, Page 780

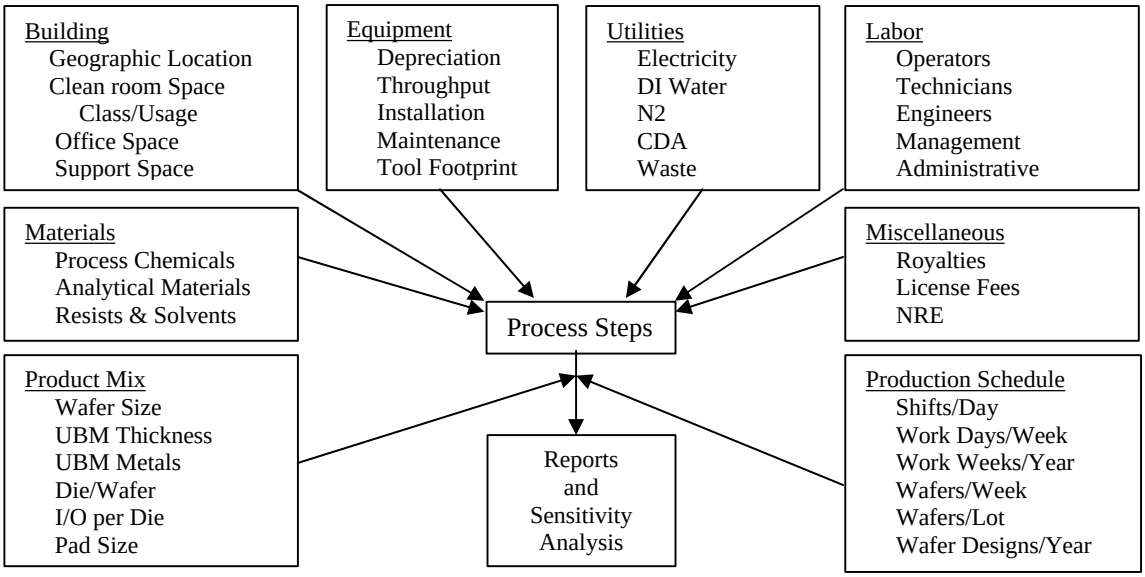


Figure 2. Cost Model Structure.

The standard approach in using this model it to create a detailed step lists for each process, identify all the resources needed for each operation, and then run several macros within the model to generate reports and sensitivity analysis. An example of a common report would be to analyze cost per wafer as a function of production volume. Examples of sensitivity analysis would be to compare the cost of each process step, or to compare the cost of the six resources. In this study we evaluated the costs of plating electroless nickel and gold on aluminum based semiconductors. Table III list the process variables which were studied.

Table III. Process Variables (shaded column represents the base case).

Shifts/day		1 shift – 5 days/wk	2 shifts – 5 days/wk	3 shifts – 7 days/wk	
Wafer Size (mm)	100	150	200	300	
e-Ni Thickness (um)		2	5	10	25
Plating Area (%)		5	10	25	50
Resist & Plasma Processes		Plasma Descum	None	Backside Resist	
Plating Line Automation		Manual	Fully Automated		
Metal Layers		Ni/Pd	Ni/Au	Ni/Pd/Au	
Pad Metallurgy		Cu	Al		
Deposition Technology		Electroplated	Electroless	Sputtered	

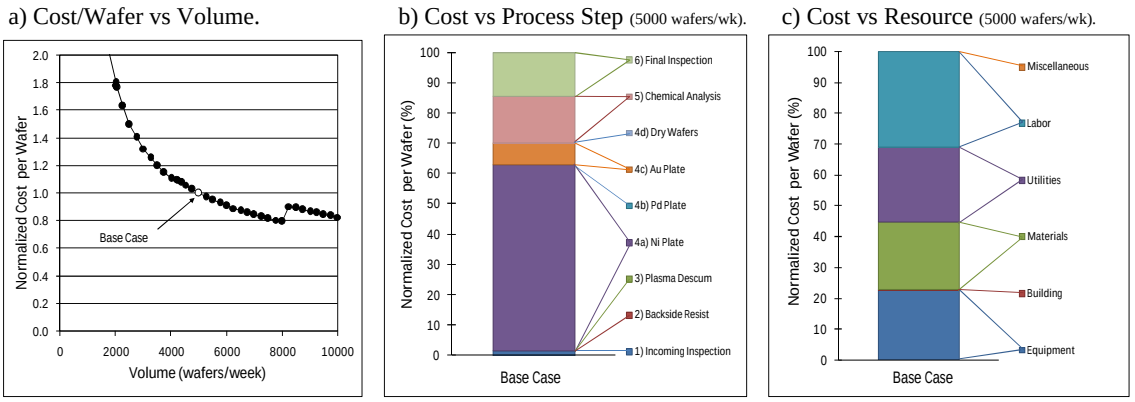


Figure 3. Base Case for Electroless Nickel/Gold Plating on Aluminum Semiconductors.

Figure 3 shows the sensitivity analyses for the electroless plating "base" case. This example is used as the standard upon which all other process variations in this paper are compared. The variables for this base care are shown in the shaded column of Table III. All of the costs have been normalized to 1 at a production volume of 5000 wafers/week (actually costs in $ are also output from cost model).

The cost per wafer is observed to decreases significantly with increased production volume. A step function in this curve is observed around 8000 wafers per week. This represents a point where a second plating line is required to plate a large number of wafers per week, at a work schedule of two shifts working 5 days/week. The Cost vs Process Step bar graph shows that the combined electroless plating operations (4a-d) represent 70% of total cost. Note that "4a Ni Plate" includes all the equipment costs (plating line), while "4b Plate Pd" and "4c Au Plate" are separated out into individual categories and a majority of their costs are associated with chemicals. The drying costs "4d Dry Wafers" in this example is zero because it is included in the fully automated equipment cost. In the Cost vs Resource bar graph, equipment, Materials, utilities, and labor encompass a majority of the costs with a fairly even distribution between them. The building and miscellaneous costs are observed to be small percentages of the total cost.

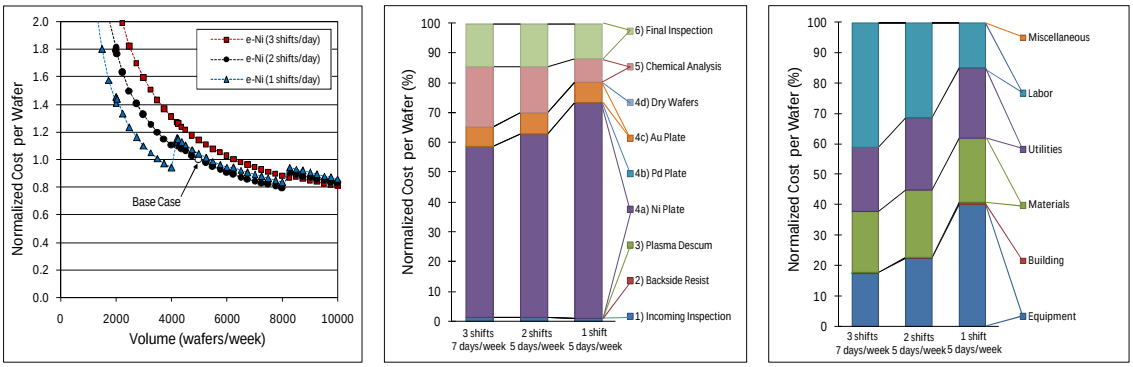


Figure 4. Shift Dependence on Ni/Au Plating Cost.

Adding a third shift to the production schedule is observed to eliminate the need for a second plating line at higher volumes. Decreasing the work schedule to one shift per week results in the need to add a third plating line in order to process more than 8000 wafers per week. At low volumes, one shift is the lowest cost scenario and at high volumes one shift is the highest cost. As the number of shifts increases, labor becomes a higher percentage of the total cost and the equipment becomes a lower percentage.

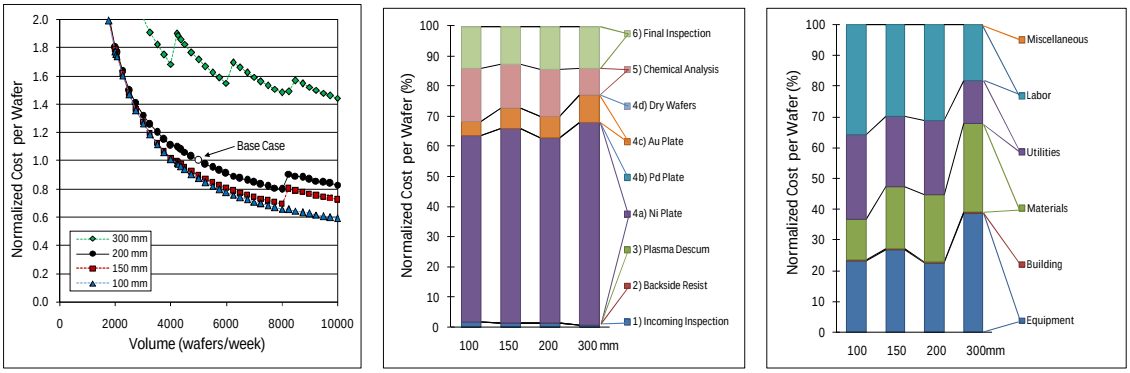


Figure 5. Wafer Size Dependence on Ni/Au Plating Cost.

At low production volumes, the cost of plating 100, 150, and 200 mm wafers are all similar. The cost of plating 300mm wafers is significantly higher due to the smaller lot size (100 wafer/lot for 100 mm, 50 wafers/lot for 150 and 200 mm, and 13 wafers/lot for 300 mm wafers). In addition, the cost for the equipment and materials in larger due to the need for a second plating line at 5000 wafers/week.

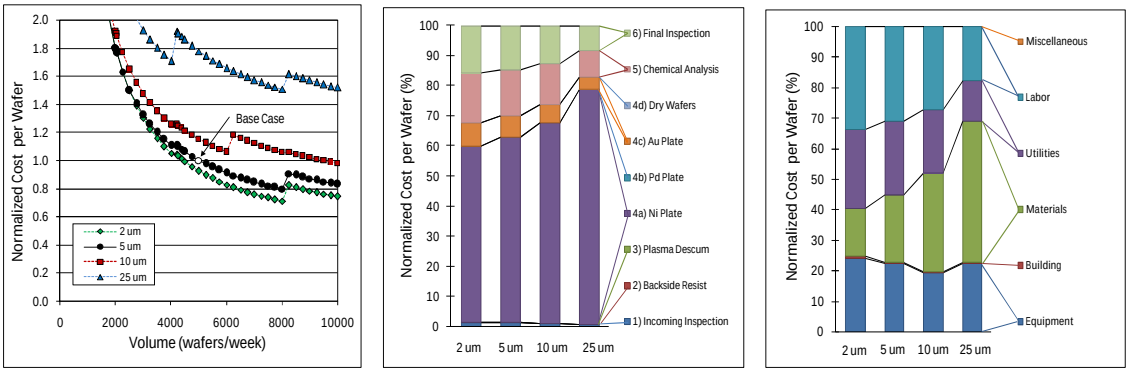


Figure 6. Nickel Thickness Dependence on Ni/Au Plating Cost.

Increasing the nickel plating thickness increases the overall cost due to: the longer times required to plate thicker films, more materials (chemicals) that are required, and higher replenishment levels. Increasing the nickel thickness also results in the requirement to add more plating lines to increase throughput.

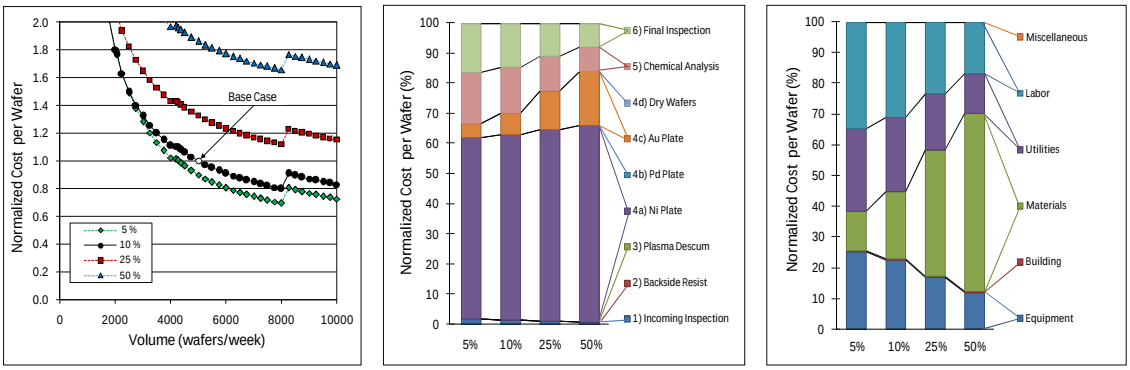


Figure 7. Plating Area Dependence on Ni/Au Plating Cost.

Some integrated circuits have significant amounts of exposed metal that will plate. Increasing this surface area has similar cost consequences to increasing the nickel thickness. A significant increase in cost is observed with increase in plating area. The higher percentage of gold chemical costs accounts for much of this increase.

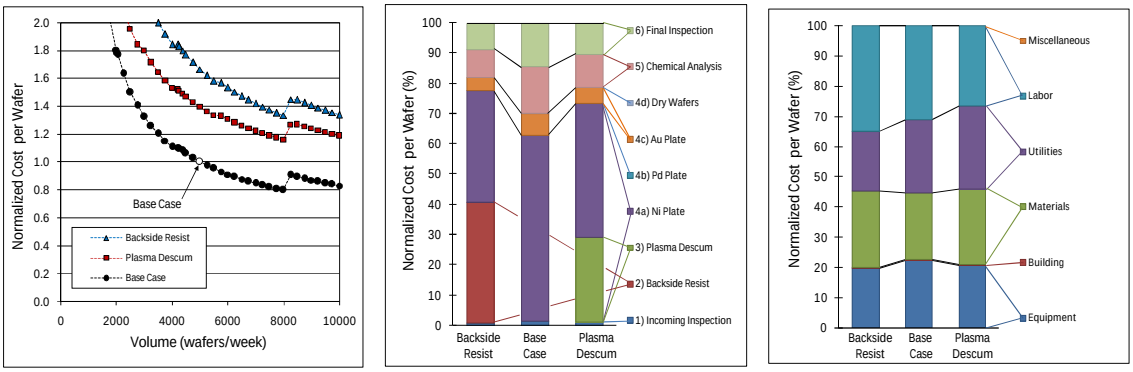


Figure 8. Backside Resist and Plasma Descum Dependence on Ni/Au Plating Cost.

The need for either plasma descum or for backside resist increases the cost of the Ni/Au plating process. A majority of the cost for backside coating is the resist cost and the additional coating equipment that is required. The resist cost is a per wafer cost and is therefore nearly independent of the number of wafers plated per week.

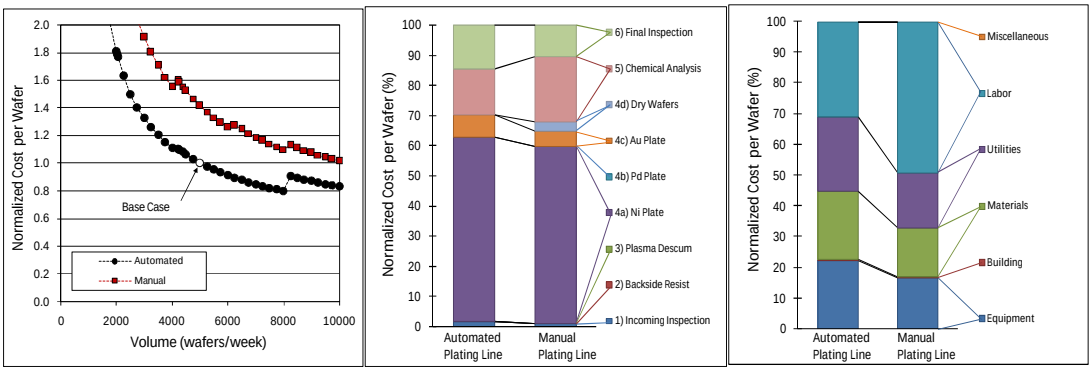


Figure 9. Equipment Automation Dependence on Ni/Au Plating Cost.

The cost per wafer is greater for a manual plating line versus the automated plating line. A majority of this cost increase is due to the increase in labor, chemical analysis, and the addition of a separate drying tool.

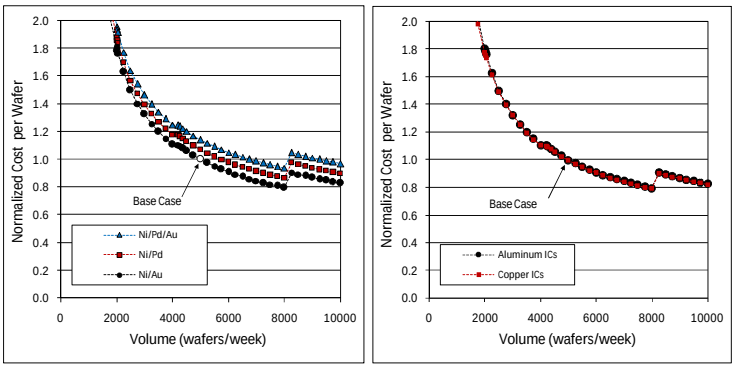


Figure 10. Palladium and Plating on Copper Based Integrated Circuits.

The use of palladium increases the cost of electroless plating. The palladium bath requires a higher level of analysis and also requires more frequent replenishment.

The cost of plating on copper based semiconductors is similar to that of plating on aluminum semi-conductors, even though the activation chemistries are different. The automated plating line in this study can plate both types of wafers.

Table IV. Process Steps for Three UBM Deposition Techniques.

	Sputtered	Electroplated	Electroless
1	Incoming Inspection	Incoming Inspection	Incoming Inspection
2	Pad Clean	Resist Coat	Nickel/Gold Plate
3	PVD	Pattern Resist	Analytical Analysis
4	Resist Coat	Descum	Final Inspection
5	Pattern Resist	Electroplate	
6	Descum	Strip Resist	
7	Etch Metal	Etch Seed Metal	
8	Strip Resist	Clean	
9	Clean	Final Inspection	
10	Final Inspection		

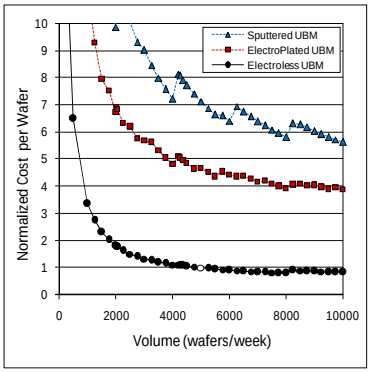


Figure 11. Ni/Au Plating Costs Compared to Electroplating and Sputtering.

Electroless plating is significantly lower in cost than electroplating or sputtering. Sputtering has significantly lower throughput and requires the addition of several high cost tool sets in order to plate 10,000 wafers/week. All three processes have high processing costs at low volumes.

References

1. T. Oppert, E. Zakel, T. Teutsch, "A Roadmap to Low-Cost Flip Chip and CSP Using Electroless Ni/Au," Proc. IEMT/IMC Symposium., Omiya, Japan, 1998.

2. A.J.G. Strandjord, S.F. Popelar, C.A. Erickson, "Low Cost Bumping for Flip Chip Applications (Electroless Nickel-Gold / Stencil Printing)", Proceedings 1999 International Symposium on Microelectronics, Chicago, IL, October 26-28, 1999, pp. 18-33.

ECS Transactions, 18 (1) 757-762 (2009)
10.1149/1.3096533 © The Electrochemical Society

Edge Protection of Temporarily Bonded Wafers during Backgrinding

Dongshun Bai[a], Xing-Fu Zhong[a], Rama Puligadda[a], Jurgen Burggraf[b], Daniel Burgstaller[b], Chris Lypka[c], and James Verzosa[c]

[a] Brewer Science, Inc., 2401 Brewer Dr., Rolla, MO 65401, USA
[b] EV Group, A 4780 Schaerding, DI Erich Thallner Strasse 1, Austria
[c] DISCO HI-TEC AMERICA, Inc., 3270 Scott Blvd., Santa Clara, CA 95054 USA

Edge chipping during backgrinding is one of the main challenges of processing temporarily bonded wafers. The edge chipping may propagate during subsequent process steps and eventually result in yield loss. We conducted a study to compare different methodologies for wafer edge protection during backgrinding, including using pre-thinned carrier wafers, large carrier wafers, edge-trimmed device wafers, and material edge modification. This paper will introduce the metrology developed to quantify edge chipping and compare the results from different protection methods.

Introduction

Industry-leading experts claim that three-dimensional (3-D) integration and packaging present the semiconductor industry's direction for the future. To meet the industry need for handling increasingly thin and fragile product wafers, Brewer Science and EV Group have collaborated to create a new temporary wafer bonding technology to provide a novel solution for processing ultrathin wafers (1-8). To prevent thin wafers from being damaged during processing, they can undergo temporary bonding to a carrier wafer by using temporary wafer bonding materials. The carrier wafer provides mechanical support during backside processing, and the temporary wafer bonding material protects the active surface of the device wafer. However, edge chipping during backgrinding is one of the main challenges for temporarily bonded wafers. The edge chipping may propagate during subsequent process steps and eventually result in yield loss. We conducted a study to compare different methodologies for wafer edge protection during backgrinding. This paper will introduce the metrology developed to quantify edge chipping and compare the results from different protection methods.

Experiment

Wafer Edge Protection Schemes

As listed in Table I, three known methods, that is, using pre-thinned carrier wafers, large carrier wafers, and edge trimmed device wafers, as well as a material edge modification method developed by Brewer Science were applied to achieve edge protection. Four wafer pairs were used in each scheme in this study. Other known methods such as edge extraction (where the device wafer edge is cut off by 1 to 2 mm during grinding) are not included in this study because of the inaccessibility of the tools.

Ex. 1016, Page 785

Table I. Schemes for Wafer Edge Protection

Scheme	Discription
Pre-thinned carrier wafers	The carrier wafers were pre-thinned to 480 m to provide more supporting area.
Large carrier wafers	Large carrier wafers were used to provide protection of the edge of device wafer
Edge trimming	The device wafers were pre-trimmed along the edge by 1 mm in width and 100 m in depth. (see Figure 1f for details)
Material egde modification	The profile of the temporary wafer bonding material along the edge of each bonded wafer pair was modified to provide good support during grinding

Materials

Wafers. Wafers used in this study were 8-inch ultra-flat silicon wafers with a diameter of 200 ± 0.2 mm, unless specified, from Silicon Quest International (SQI). The pre-thinned silicon carrier wafers with a diameter of 200 ± 0.2 mm were thinned on a DISCO DFG8540 grinder. The large silicon carrier wafers were from MEMC and had a diameter of 200.1 ± 0.1 mm. The device wafers that were bonded to the large carrier wafers had a diameter of 200 ± 0.1 mm. The edge-trimming process was done on a DISCO DFD6361 Fully Automatic Dicing Saw, and the front side of each wafer had 100 m in depth trimmed from the outermost 1-mm-wide band. The thickness of the wafers was 725 ± 25 m except for the pre-thinned carrier wafers. The wafer to be ground was called the "device wafer," and the wafer used to support the thinned wafer was called the "carrier wafer." All wafers used in this study were blank wafers having no features.

Temporary Wafer Bonding Material. The temporary wafer bonding material used was WaferBOND™ HT.10.10 material from Brewer Science, Inc.

Coating and Bonding of Wafers

All wafers were coated and bonded on an EVG® 850TB automatic bonding system at EVG USA except for the wafers bonded to large carriers, which were coated and bonded on the auto-centering system at EVG Austria. The process used for coating and bonding the wafers was the same for all wafers and is described below.

Coating and Baking. The device wafers were coated with WaferBOND™ HT.10.10 material at 1200 rpm for 30 seconds. After coating, the wafers were proximity baked at 160°C for 4 minutes. The coating thickness of one wafer was measured on a Filmetrics thickness measuring system and the coating had a mean thickness of 19.6 m and minimum and maximum thicknesses of 19.1 and 20.2 m, respectively.

Bonding. The device wafers were then bonded to the carrier wafers at 180°C and 3500 N for 2 minutes under vacuum. The bonding quality was very good, as no voids or

sharp color variations were observed by scanning acoustic microscopy (SAM) on all bonded wafer pairs.

Material Edge Modification

Material edge modification was performed to repair the non-uniformity of the bonding materials created during the bonding process. After bonding, the temporary wafer bonding materials between two bonded wafers were removed along the edge by side rinsing with solvent by 0.5 mm into the wafer edge. Then the space that was occupied by the removed material was refilled with WaferBOND™ HT.10.10 material with a nozzle along the bonded wafers' edge. The wafer pairs were then baked at 150°C for 2 minutes. After modification, the wafer edge section was inspected with a camera. The material covered the entire wafer edge including the bevel evenly.

Grinding and Polishing

Grinding was performed on a DISCO DFG8540 Fully Automatic Grinder. All wafer pairs were ground from the backside of the device wafers to about 92 m with wheel Z1, and then further ground to about 52 m with Poligrind wheel Z2. To reduce the stress generated during grinding, a dry polishing step was used to smooth the ground wafer surface on a DISCO DFG8140 Fully Automatic Polisher. The final thickness of the ground wafers after dry polishing was about 50 m. Detailed information on the wheel type used in each step is listed in Table II.

Table II. Detailed information on grinding and dry polishing wheels

Wheel	Z1	Z2	Dry Polishing Wheel
Code	DRSJ0072	DRAA0005	PAZZ0014
Specification	GF01-SDC320-BT300-50	PW-005	DPW-014 DP-F05
Size	200x4Wx5T-SD	200x7Tx3W	300x11Tx60

Wafer Edge Inspection

After grinding and dry polishing, the device wafer edges were inspected with a Keyence VHZ100 microscope. A quick survey was conducted first around the wafer circumference to ensure that no large chips were missed during inspection. To quantify the edge chipping, the wafer edge was inspected every 30 degrees along the edge with 100X magnification. The number of chips within each range was counted and categorized into four grades according to the chipping distance into the wafer edge. The four grades are <10 m, 10 to 50 m, 51 to 100 m, and 101 to 200 m.

Results and Discussion

For all 20 wafer pairs that went through the grinding and dry polishing process, we noticed only a single small blister on each of two device wafers. We did not observe any cracking or delamination along the wafer edges. The largest chipping we observed with a microscope was less than 150 m into the wafer edge. Thus, the temporary wafer bonding material worked very well during backside grinding and polishing of the device wafers.

Ex. 1016, Page 787

Because we previously observed large chips and cracks in the thinned device wafers while they were still bonded to the carriers after shipping, we decided to do the edge inspection immediately after dry polishing to eliminate the chance of later detecting possible shipping damage to the wafer edges. Automatic wafer edge inspection systems are currently on the market, such as the VisEdge system from KLA-Tencor and the E30™ system from Rudolph Technologies. However, we could not find a place that could do both the grinding and then the automatic wafer edge inspection. Therefore, we developed a metrology to quantify the wafer edge chipping after grinding and stress relief.

As discussed above in the Experiment section, we performed a 12-point inspection along the wafer edges on each of the thinned device wafers and categorized the chipping according to the distance into the wafer edge. The overall results are listed in Table III.

Table III. The number of chips and their size distribution on 12 inspection spots

Scheme	**Chipping numbers on twelve inspected spots**				
	Total number	<10 m	10 to 50 m	51 to 100 m	101 to 200 m
Pre-thinned carrier wafers	16±5.6	9.0±3.7	6.3±3.2	0.8±1.0	0.0±0.0
Large carrier wafers	39.3±8.7	18.8±2.5	19.3±7.5	1.3±1.3	0.0±0.0
Edge trimming	7.3±2.1	5.5±0.6	1.8±1.7	0.0±0.0	0.0±0.0
Material edge modification	25.3±8.5	9.3±5.4	15.5±2.9	0.5±1.0	0.0±0.0
Control	66.3±12.1	14.0±5.5	43.3±8.5	8.0±9.5	1.0±1.2

All four edge protection methods showed improvement in reducing the edge chipping. Edge chipping was reduced by at least 40% based on the total chipping number. The chipping size was remarkably reduced as well, and we did not observe chips larger than 100 m for all protection schemes. Figure 1 shows pictures of the device wafers taken by the microscope after backgrinding and dry polishing. These pictures were taken of the same position on the thinned wafers for different schemes. The grid lines in the picture were used for determining the chipping size into the edge.

Among the four protection schemes, the edge trimming method provided the best edge protection. It not only reduced the total number of chips by 89% but also reduced the chip size. All chips from this scheme were less than 50 m and most of them were less than 10 m. Figure 1f shows a picture of the edge-trimmed device wafer. By cutting the wafer along the edge by a width of 1 mm and a depth of 100 m, the resulting wafer edge was protected very well during grinding and polishing to reduce edge chipping.

The pre-thinned carrier wafers provided the second-best results. The total number of chips was reduced by 76 percent and the chip size was dramatically reduced as well. The pre-thinned carrier also provided good thickness uniformity so that the total thickness variation of the device wafer after thinning could be reduced as well.

The material edge modification method also exhibited remarkable improvement in reducing chipping. The uniform support from the materials around the edge of the bonded wafer pair provides a better support of the device wafer edge.

The protection from the large carrier wafer is not as good as that of the other methods. The diameter of large carrier wafers in this study was 200.1 ± 0.1 mm, which may not cover the edges of device wafers with a diameter of 200 ± 0.1 mm, and the alignment becomes very critical in this situation. A large carrier wafer with a diameter of 200.2 ± 0 mm is ideal in this situation and is also acceptable for most semiconductor equipment.

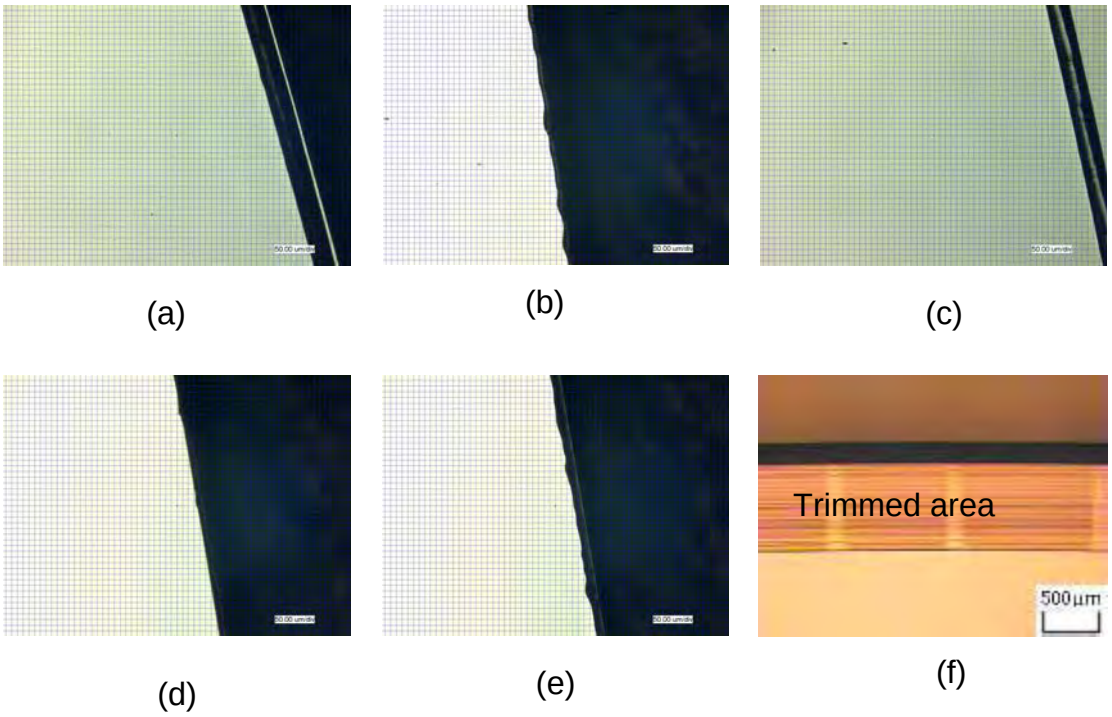


Figure 1. Pictures of device wafer edges after grinding and dry polishing for schemes of a) a pre-thinned carrier wafer, b) a large carrier wafer, c) edge trimming, d) material edge modification, e) control. f) shows a front view of the edge-trimmed wafer.

Statistical analysis of the chipping data

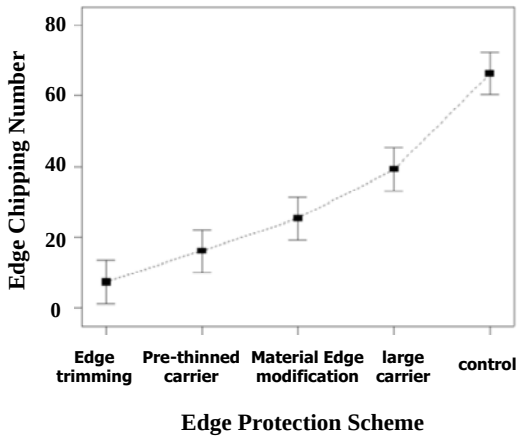


Figure 2. Statistical analysis of the chipping data based on total number of chips from 12 inspection spots.

We also analyzed the total number of chips from the 12 inspection spots by using Design-Expert software. The study was treated as one factor 5 level designed experiment with four replicates. The analysis of variants (ANOVA) resulted in a model F-value of 32.34, which implies that the model is significant. There is only a 0.01% chance that a model F-value this large could occur due to noise. The 95% confidence intervals for each scheme are also plotted in Figure 2. The model graph clearly shows that the effect of treatment is very significant, with edge trimming giving the least chipping, followed by pre-thinned carrier wafers, material edge modification, large carrier wafers and control wafers without treatment.

Conclusions

Four edge protection schemes were tested to protect wafer edges from chipping during backgrinding. All schemes showed improvement in reducing the number of edge chips. The edge trimming method is the most effective method of the four, followed by the pre-thinned carrier wafer and the material edge modification methods. Large carrier wafers used in this study may not be able to protect the entire device wafer edge, they only reduced the number of chips by 40%.

Acknowledgments

Thanks to Mark Privett from Brewer Science for his suggestions on this study. Also, the material edge modification assistance provided by Gary Brand and Jeremy McCutcheon from Brewer Science is greatly appreciated. Thanks to Chad Brubaker from EV Group for his support on wafer coating and bonding.

References

1. J. Lamb, B. Kim, S. Pargfrieder, *Solid State Technology*, Vol. 51, pp. 60-65 (2008).
2. P. Singer, "Through-Silicon Vias: Ready for Volume Manufacturing?" *Semiconductor International*, No. 3, online (2008).
3. J. Vardaman, "3-D Through-Silicon Vias Become a Reality," *Semiconductor International*, Vol. 30, No. 6, online, (2007).
4. H. Kirchberger, B. Kim, S. Pargfrieder, *Advanced Packaging*, pp. 45-49, March (2008).
5. R. Puligadda, D. Edwards, *Future Fab International*, Vol. 24, pp. 107-110, (2008).
6. D. Bai, W. Hong, J. Dachsteiner, A. Jouve, R. Puligadda, C. Brubaker, T. Tian, *IMAPS 2008: Proceedings of the IMAPS 40th International Symposium on Microelectronics*, pp. 222-227, (2008).
7. W. Hong, R. Puligadda, A. Smith, D. Bai, S. Pargfrieder, C. Brubaker, S. Pfeiffer, *IMAPS 2007: Proceedings of the IMAPS 40th International Symposium on Microelectronics*, pp. 1035-1040, (2007).
8. R. Puligadda, S. Pillalamarri, W. Hong, C. Brubaker, M. Wimplinger, S. Pargfrieder, *Materials Research Society Symposium Proceedings*, Vol. 970, paper no. 0970-Y04-09, (2007).

ECS Transactions, 18 (1) 763-768 (2009)
10.1149/1.3096534 © The Electrochemical Society

Wafer Level SAW RF Filter Packaging With Through Wafer Via Interconnection

Young Do Kweon, Seung Wook Park , Tae Hoon Kim, Ju Pyo Hong, , Si Joong Yang, Job Ha, Tae Ho.Kim*, and Sung Yi

[1]PKG Team, Corporate R&D Institute, Samsung Electro-Mechanics Co., LTD
314, Meatan 3-Dong, Yeoungtong-Gu, Suwon, Gyunggi-Do, Korea
[2]Electroplating Group, Manufacturing Engineering R&D Institute, Samsung Electro-Mechanics Co., LTD

This study describes the wafer level 0.8x0.6 mm^2 SAW filter (Surface Acoustic Wave Filter) with interconnection via and LiTaO3 (LT)-LiTaO3 (LT) wafer bonding structure. The via formation for interconnection is based on smaller package manufacture. The interconnection via is made by blind via and wafer thinning process. The blind via is formed by sand blasting process or laser drilling process. The eutectic bonding is applied to LT and LT bonding with Au-Sn for device protection and pads interconnection. These processes enable the wafer level saw filter. The result of electrical performance and laser process and sandblaster process for wafer level SAW filter package is discussed

Introduction

Recently, the mobile cellular phone market has strongly requested the small size and slim type mobile phone by rapid technology development. So the SAW filter device which is a main component of mobile phone parts needs to decrease the size and increase the performance for adjusting market demand. The conventional SAW filter package has used a SMD (Surface Mounted Device) and Mold type CSP (Chip Scale Package) package using ceramic substrate which is to protect the package from outside environment. But it's difficult to make the smaller size package because they need space for chip SMT (Surface Mount Technology) and encapsulation. [1-6]

WLP (Wafer Level Package) is prospected on the smaller size, lower height, less weight and higher performance for the SAW filter. To realize the WLP SAW filter, wafer level technology has to be studied on wafer bonding, wafer thinning and via drilling. In case of wafer bonding, the Au/Sn eutectic bonding technology has been selected with low temperature and no fume during bonding process in our previous study. The thinning technologies which are lapping and grinding supply the low profile and low weight on the SAW filter. [1, 4] However, the drilling process has an extreme resistance by standard drilling techniques like RIE and wet- etching. Our study focus on the interconnection by via drilling processes which are laser drilling and sand blasting. Two methods are promised on an interconnection drilling process for LT Wafer. In case of sand blasting process can easily produce the via hole but have a limitation for via size control and surface damage of via inside. Laser drilling can make smaller size and fine surface of via inside that also generated 2nd troubles like debris and HAZ (Heat Affected Zone). [3]

In this study, we have reviewed the bonding, thinning and interconnection process for wafer level SAW filter. And we report on reliability and electric performance of the 0.8x0.6 mm^2 WLP SAW Filter.

Ex. 1016, Page 791

Wafer Level Saw Filter Design and Process

Wafer Level Saw Filter Design & Process

The figure 1 is 0.8x0.6 mm^2 wafer level SAW filter design. This package consists of a package wafer which contains through via by drilling process and a SAW filter device wafer with hermetic metal seal layers. [1, 2, 3]

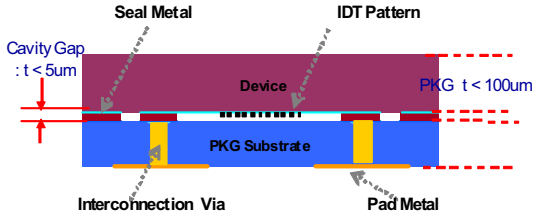


Figure 1 Wafer level SAW filter package

Figure 2 illustrates the wafer level process for 0.8x0.6 mm^2 SAW filter. The blind via can be formatted using UV, Green laser drilling and Sand blasting. The interconnection via is generated by via filling and thinning and bonding process. The cavity for acoustic wave is made by Au-Sn bonding layer. [3]

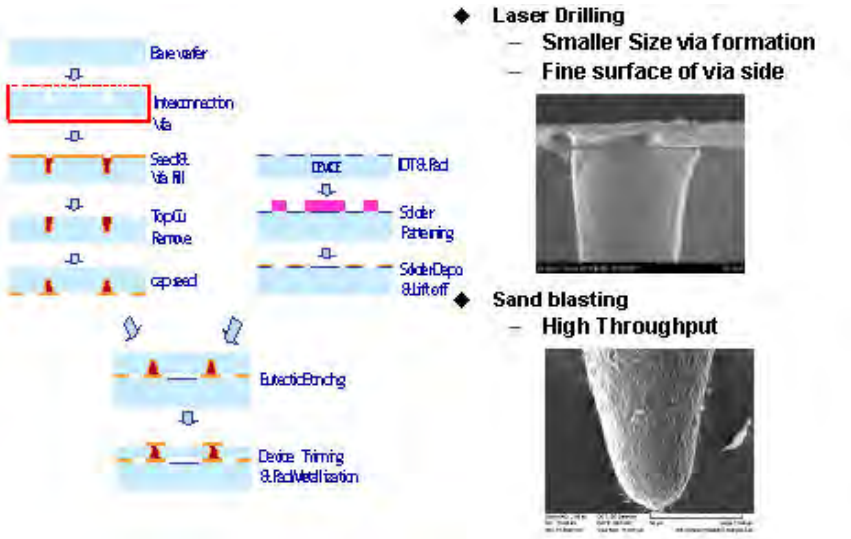


Figure 2 Process flow for wafer level SAW filter

Wafer level SAW filter Design and Concept

To confirm the stress concentration point, we try to do FEM simulation during bonding profile which apply to max temperature of 280 ℃. Figure 3 shows the stress distribution of SAW filter package after wafer bonding process. To show the stress status clearly, each components of the package are separated from the others. There are stress concentration areas in each component of the package. After reliability test, the failure modes also confirm it by SEM analysis in figure 3.The maximum stress area is via side. The mismatch of the CTE between these two components will cause higher thermal stresses

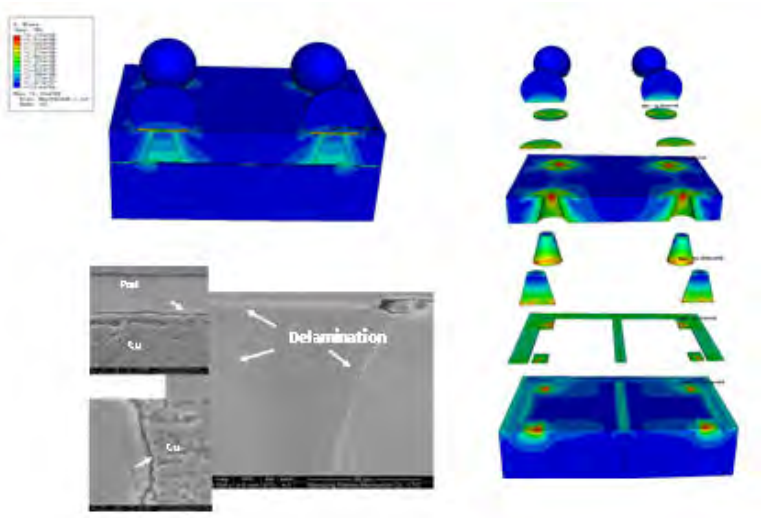


Figure 3 Stress distribution of the 0.8x0.6 mm^2 cap wafer

Via formation

Sand Via formation The sand blasting is a general method for MEMS device that can also support a via drilling. In case of LT wafer, it's difficult to apply a processing such as RIE (Reactive Ion Etching) and chemical etching because it's stable to chemical reaction and weak for thermal stress by CTE mismatch. So Sand blasting is a promising for via drilling of LT wafer. To get a small and deep blind via, it's important to make a DFR (Dry Film Resistor) patterning and apply pressure and process time for a sand blasting. Our study gets a 130um via size and 120 via depth by sand blasting. Figure 4 is sand blasting process flow. Figure 5 show result for sand blasting process which of a) is inside surface and via shape and b) is filling result by electroplating.

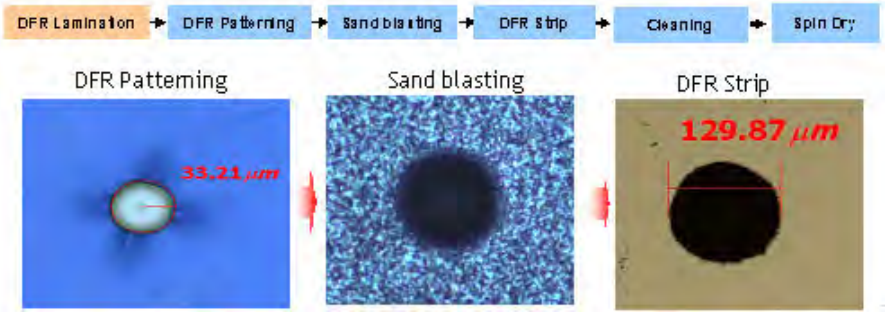
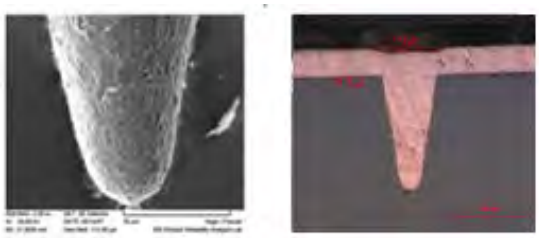


Figure 4 Sand blasting process flows



a)Via surface b) Via filling

Figure 5 Blind via filling by Electroplating

Laser via formation The laser drilling process is a next generation method for via formation. It produces blind via with a smaller size and deeper depth within very short time. However it make the 2nd troubles like HAZ and debris which disturb the via filling and adhesion trouble. We apply a etching and boiling method to clear these troubles. Blew Figure 6 is step-up process for clearing HAZ and debris. [3, 4]

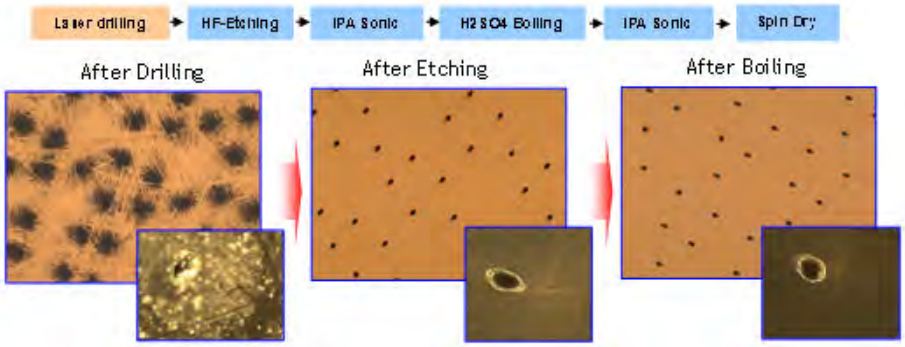
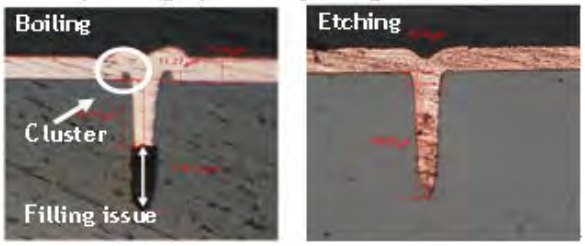


Figure 6 Laser via process with cleaning steps

Figure 7 illustrate the before and after applying via cleaning steps. After etching process the blind via was fully filling by electroplating. [8, 9]



a) Before etching b) After etching

Figure 7 The electroplating result after applying etching

Figure 8 is the Cu removing processes which is for making blind via to interconnection and patterning bonding and sealing pattern. So the flatness control is a key issue. The maximum roughness is about 38nm after Cu removing.



Figure 8 The electroplating result after applying etching

Wafer Bonding Process

The bonding process is an important to get package reliability and an electrical function for SAW filter device. A contamination can make an electrical performance shift and non-wet bonding produces a reliability failure. This study applies an Au-Sn Eutectic bonding which is no-fume and high stable after reflow process for SMT. Figure 9 is

Ex. 1016, Page 794

bonding result for pressure deviation. The high pressure bonding shows a high shear strength which was avg.1.44 kgf without void and no-wet areas. But Low-pressure bonding was avg. 0.77 kgf with void and non-wet area.

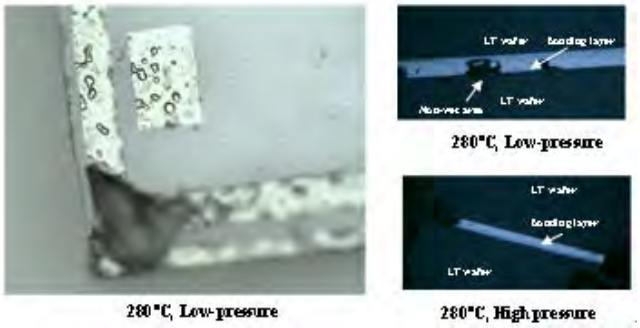


Figure 9 The wafer bonding result

LT Wafer Thinning

The Wafer thinning process is a main process for making through via. It's applied after wafer bonding process. That time via is blind so we have to open the cu filling area of the package wafer by back side grinding and polishing. And it controls the package profile through the device wafer thinning. Figure 10 shows a wafer thinning procedure and results.

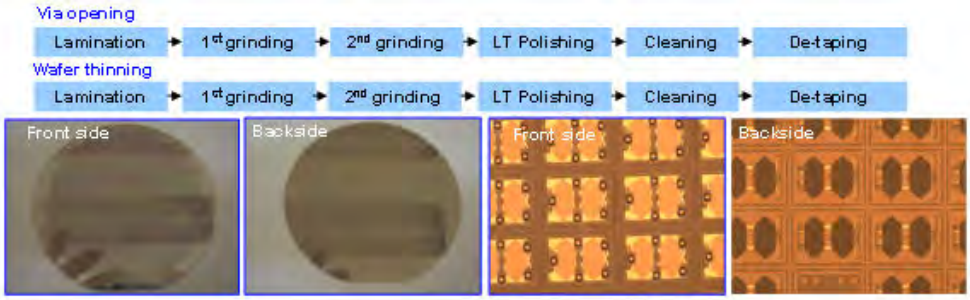
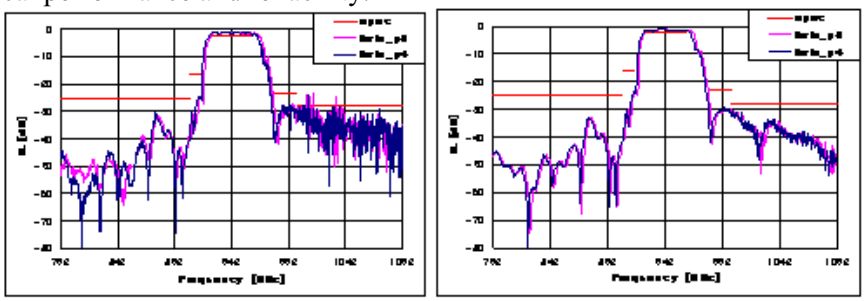


Figure 10 The wafer thinning result

Result

The reliability test result, this package achieved the good result, the MSL L1, PCT 96hr, TC 135Cycle (150 ℃~ -65 ℃/30min) was passed. Figure 11 shows the frequency response of the wafer level SAW filter after completing process that adapted the laser via drilling and Sand blasting. That satisfied the specification of electrical performance and reliability.



a)Sand Blaster via b)Laser via

Figure 11 Measurement result of frequency response

Conclusions

In this paper, we describe the 0.8x0.6x0.25mm^3 wafer level SAW filter by sand blasting and laser drilling. The through via for interconnection was formatted by both drilling process and electroplating and wafer thinning. These processes show a good workability for through via formation. The wafer level SAW filter by the laser via process and sand blasting shows a good frequency response and reliability. We believe that this laser via drilling process and sand blasting can provide the higher yield and performance and more compact SAW filter applications

References

1. Tae Hoon Kim and Won Kyu Jeung, "Miniaturization and Optimization of RF SAW Filter Using Wafer Level Packaging Technology" Electronic Components and Technology Conference, 2007 pp. 574-579
2. Gao Shan and Ju-Pyo.Hong "Study of a Wafer Level Package (WLP) for Surface Acoustic Wave (SAW) Filter" EMAP(2006)
3. Seung Wook Park and Ju Pyo Hong, "Fabrication and Optimization of Wafer Level SAW Filter Package Using Laser Via Drilling" ESTC (2008)
4. P. Selmeier, R. Grunwald, A. Przadka, H. Kruger, G. Feiertag and C. Ruppel, "Recent Advances in SAW Packaging" , IEEE Ultrasonic Symposium, 2001, pp 283-292.
5. Hiromi Yatsuda, Taira Horishima, Takeshi Eimura, and Takao Ooiwa, "Miniaturized SAW filter Using a Flip- Chip Technique", IEEE Ultrasonic Symposium, Vol. 43, No. 1, Jan. 1996, pp 125-130.
6. H. Meier, T. Baier, and G. Riha, "Miniaturization and advanced functionalities of SAW devices" Proceedings of the IEEE Ultrasonic Symposium, 2000, pp. 395-406.
7. M. Goetz and C. Jones, "Chip Scale Packaging Techniques for RF SAW Devices" IEEE International Electronics Manufacturing Technology (IEMT) Symposium, 2001, pp. 63-66.
8. S. Mailis, G.W. Ross, L. Reekie, J.A. Abernethy and R.W. Eason, " Fabrication of surface relief gratings on lithium niobate by combined UV laser and wet etching" ELECTRONICS LETTERS 12th October 2000 Vol. 36 No.21 pp1801-1803
9. Mahn-Jick Lim "Method of Etching a surface of a substrate comprising LiTaO3 and Chemically similar materials" United States Patent, Re.29336 (1977)

ECS Transactions, 18 (1) 769-776 (2009)
10.1149/1.3096535 © The Electrochemical Society

High Volume, Low Cost Multi Die Packaging Using Conductive Polymer Interconnect

M. E. Robinson[a] and J.S. Leal[a]

[a]Vertical Circuits, Inc., Scotts Valley, California, 95066, USA

Wire bonding multiple die in a single package, PoP (Package on Package) and, more recently, TSV (thru silicon via) approaches are often considered for small, portable consumer electronic devices. Conductive polymer interconnect is a new approach that provides cost and size advantages for a wide variety of applications. Conductive filled polymer materials are well suited for chip to package and chip to chip interconnect. They are compatible with existing equipment and material sets used in a typical semiconductor assembly facility, and capital costs for adding this capability are relatively low. This method also offers size and cost advantages for single die packaging of image sensors and other die, where there is a requirement to connect to the front side of a die from the backside, without occluding the front surface. Wafer level packaging (WLP) methods and array processing greatly reduce costs, making the technology competitive with traditional wire bond techniques.

Introduction

High volume portable and consumer electronic devices are driving technology and product developments in many areas where cost, reliability, and functionality are of primary importance. In a portable device, minimizing the footprint, height, and weight of electronic components is of critical importance.(1) To address these demands, we have developed a side-interconnect ESV (edge silicon via) die stacking technology that integrates easily into a typical mainstream assembly factory. Conductors are applied as a liquid over the edges of a stack of integrated circuit die and after cure the conductors function as virtual vias, electrically interconnecting the die in the stack with each other and with a substrate, or circuit board upon which the die are mounted.

Conformal conductors offer size, performance and cost advantages that improve upon conventional wire-bond assembly structures.(2) For this work, we constructed 16Gbyte 8-high, 0.8mm µSD flash memory cards, and a 64GByte Solid State Drive. The method is also suitable for high density DRAM applications, interconnecting stacked packages as an alternative to POP, and for packaging single die which require connection to the front surface from the back, such as CMOS image sensors.

In our earlier work we focused on extrusion dispense of conductive adhesives for edge aligned die stacks.(3) For this effort, to achieve more precise control, finer conductor diameters, and tighter pitches, we introduce the use of jetting to dispense the conductors.

Process Overview

Parylene is utilized as a conformal coating and edge insulation on the die so that a conformal conductor can be formed on the edge of a die stack without electrically shorting to the silicon die edges. Parylene films as thin as 1 micron exhibit dielectric

Ex. 1016, Page 797

strengths greater than 100 volts, and therefore exceed the requirements of most portable microelectronic systems, which typically operate at potentials well below 5 volts.(4) A conducting silver filled liquid is applied to the edge of a die stack to form inter-die connections, after which the stack is subjected to a low temperature cure to solidify the material into a conductive glass-like inter-die electrical connection. HVM manufacturing techniques compatible with dice before grind, and dice after grind process flows allow the sidewalls of the die to be coated with a thin layer of parylene while still in wafer form, or arrayed on a substrate strip. As the polymer insulation process coats all surfaces of the die, laser ablation is used to remove the polymer from the die pads for electrical connection. As the die are maintained in an array, the pre-stacking die preparation steps benefit from the same economies of scale normally enjoyed by wafer foundry processes. If the die have not been designed for edge connection stacking, or if the die pads are not in the proper location, an optional RDL layer can be applied to reposition the pads prior to the die insulation step.

Utilizing mainstream equipment typically found in semiconductor assembly areas for dispensing liquid epoxies, hi speed jetting techniques are used to apply the inter-die conductors to the sides of the die stacks. Figure 1, below, shows a 16GB memory card formed by stacking 8 die insulated and interconnected in this manner.

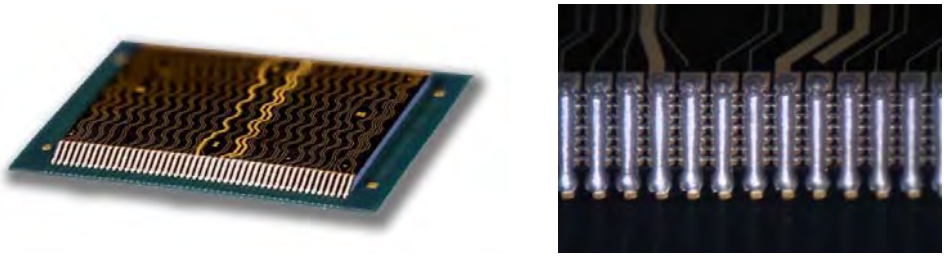


Figure 1: Inter-Die Edge Connections

Inter-die connections in contact with the edge of the die result in small component foot prints barely larger than the die. The interconnects are as close as you can get to TSV, without the need to make any vias in the die, and can be used to interconnect stacks of leadless packages, or to package single die on a substrate or carrier.

Following formation of the conformal edge interconnections, substrate strips with attached and interconnected stacks of die are molded using standard IC package molding equipment and methods to create any one of many common memory card formats currently popular. After singulation, polymer interconnected packages cannot be readily distinguished from wire-bonded devices by external inspection, apart from a smaller form-factor for a given internal die size. Die which are too big to stack in standard format memory cards with wire-bond methods will fit the required footprint when conformal interconnect is used. Substantial cost reductions are realized with an array-based process due to the elimination of repeated die attach and pad-based wirebond processing steps, and the reduction or elimination of gold.

Process Details

Semiconductor wafers are thinned and singulated using standard, back-end thinning techniques. The process is compatible with dice before grind as well as dice after grind manufacturing flows. Following thinning, application of a die attach film (DAF) and

Ex. 1016, Page 798

singulation, the die are stacked and laminated to each other in an array atop an appropriate substrate for the intended application. In some applications the die are stacked with a small translational offset in order to expose the connection pads, while in other cases a spacer can be used between the die to allow the conformal conductor access for electrical connection. The vertical conductors are applied as a liquid to the side of the stack using standard fluid dispense equipment and hybrid jetting dispense techniques, and are cured at relatively low temperatures between 100°C and 200°C. The completed stack can be electrically and physically attached to any substrate, leadframe, or printed circuit board using the same material and equipment used to form the vertical interconnects.

Figures 2 & 3 detail the process flow for 8-die flash memory stacks. To lower the cost for flash applications, the edge insulation step is performed after the die have been stacked in a tier, or terrace, allowing much thinner structures and enabling a sequential process flow with no repeating steps. The elimination of wire bonding from the process allowed us to minimize the offset to as little as 110um, as only a small portion of the die pad need be exposed to make a good electrical contact to the vertical interconnect. Required offsets are less than half that required for wire-bond stacking, as there is no need to provide clearance for the wire bonding tool, and result in modules that are compliant with both JEDEC Component and SDI Memory Card physical and reliability standards.

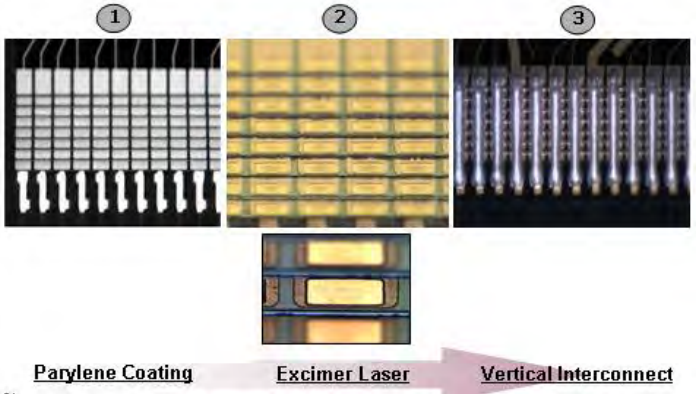


Figure 2: Stacking, Edge Insulation, and Vertical Interconnect

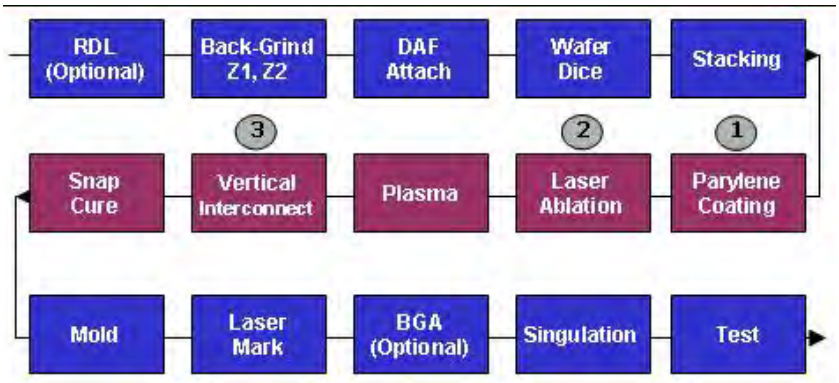


Figure 3: Dice after Grind, Coat after Stack Process Flow

Conformal Interconnects

Wirebonding to extremely thin die presents a significant challenge. A process that reduces or eliminates the applied forces on the die during wire-bonding results in both enhanced yields and superior product reliability over traditional electrical interconnect methodologies. In this new method, conformal polymer conductors are formed by use of high speed hybrid jetting techniques. The process can be implemented with standard jetting equipment currently used for high volume electronic assembly. Productivity rates for both vertically aligned and tier stacked devices exceed 750 interconnects per minute for 8 high die stacks . Productivity rates depend on the length of the interconnect, not the number of connections made, and can be independent of the number of die in the stack. As such, specifying a thinner die and/or smaller offset allows us to stack more die at the same rate per stack. Figure 4 shows conformal die interconnect in the process of application to the side of an 8-die flash stack on a µSD memory card substrate.

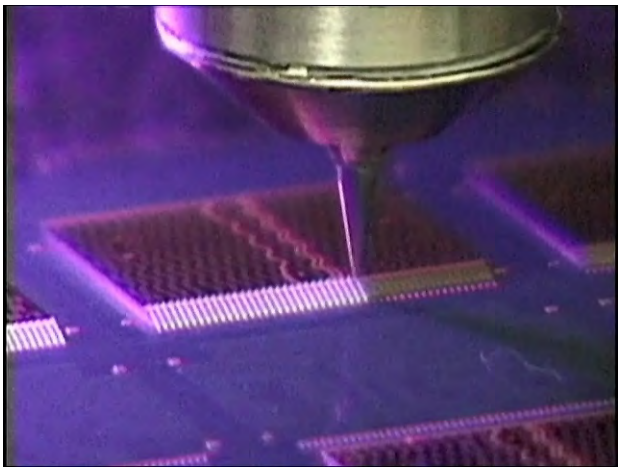


Figure 4: Hybrid Jetting Application of Conformal Vertical Interconnects

A variety of conductive paste materials can be used, the majority of which are composed of epoxy binder resins and are highly filled (>80% w/w) with silver (Ag) particles. Absolute maximum particle size is specified not to exceed 50µm in order to prevent needle clogging, obtain fine line widths, and minimize pitch between lines. Agglomerations of silver particles must be minimized as well. Paste viscosity can vary within the range between 20-300KcP. A high thixotropic index is preferred. Even though materials with a thixotropic index between 2.0 and 3.5 have been used successfully, materials with an index between 5.0 and 6.0 are preferred.

The dispense valve or pump must be capable of providing nanoliter dispense volumes at the needle tip with picoliter adjustability and rapid start/stop cycles. Commonly used needle orifices are between 75µm and 125µm. Dimensional precision of mechanical automation and machine vision systems used should be better than +/-25µm. Gang deposition (concurrent formation of multiple interconnects), stencil printing, and other approaches can be utilized as well. We have had the best success in controlling the size and shape of interconnects, and meeting high volume throughput requirements with the use of a hybrid dispense method that combines the volumetric accuracy of jetting with the positional accuracy and material control of needle dispensing.

Specifically designed cure cycles were closely followed, and kept well below 180°C to minimize thermal effects on the memory die. Material cure schedules were characterized and optimized appropriately to control key material physical and electrical characteristics that affect the performance and reliability of the interconnects. Material properties can be modified for specific package designs to manage interactions between the paste, deposition equipment, and substrate choices. Material wetting, electrical conductivity and interfacial adhesion are attributes that can be modified through control and balance of parameters during the cure cycle. Figure 5 demonstrates the dramatic difference in observed failure mode for the same material that resulted from modifying the cure cycle; the sample on the left exhibiting a cohesive failure, whereby the failure occurs within the material, while the sample on the right exhibiting adhesive failure, whereby the failure occurs at the interface between the conductor and the die pad. .

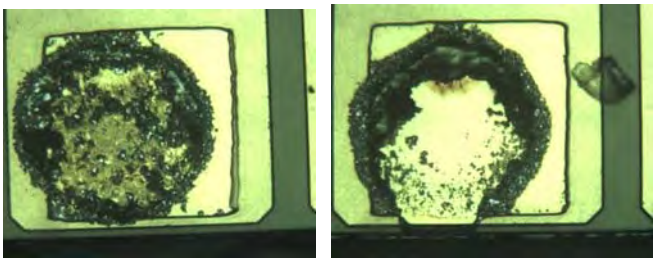


Figure 5: Shear Test Failures – Cohesive on Left; Adhesive on Right

Figure 6 shows the measured difference in adhesion we observed under 6 different cure conditions and illustrates the wide range of results that can be achieved by modifying the cure cycle.

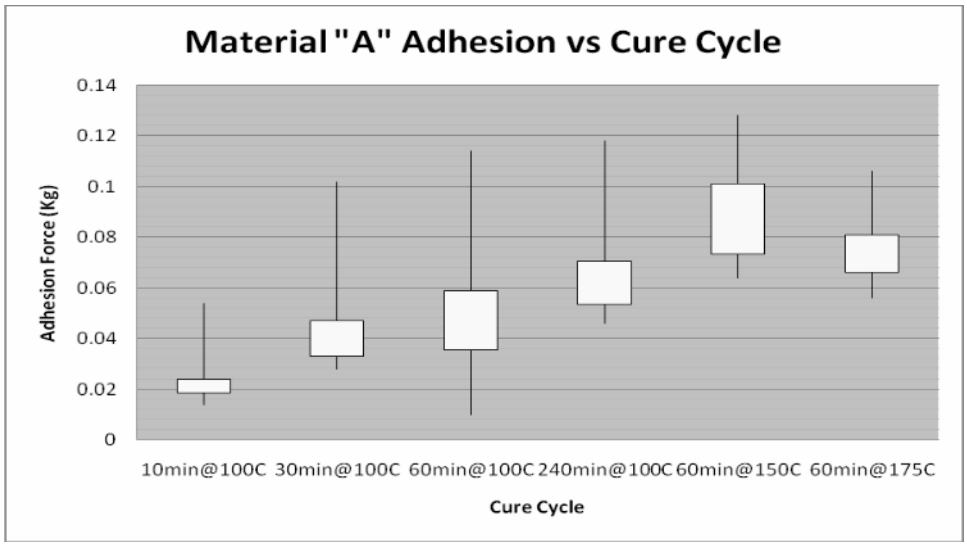


Figure 6: Measured Adhesion vs Cure Cycle

Electrical conductivity can also be modified and controlled through selection of curing profile. Figure 7 shows the measured resistance for sample groups (3 samples each) of Material "A" subjected to the same curing schedules shown in Figure 6. It is important to note that the curing schedule that results in the best adhesion is not the one that provides the lowest resistance, illustrating the need for to design the process for an optimum balance between material adhesion and electrical conductivity for the target application.

Ex. 1016, Page 801

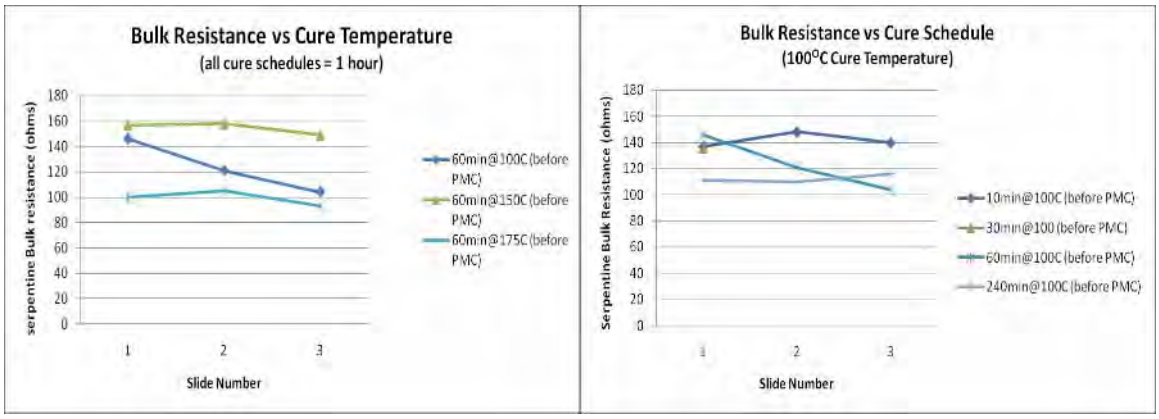


Figure 7: Cure Cycle vs. Resistance

Figure 8 shows a cross-sectional view of a typical conformal conductor on a flash memory tier stack after optimization of the sequential dispense process and the cure process.

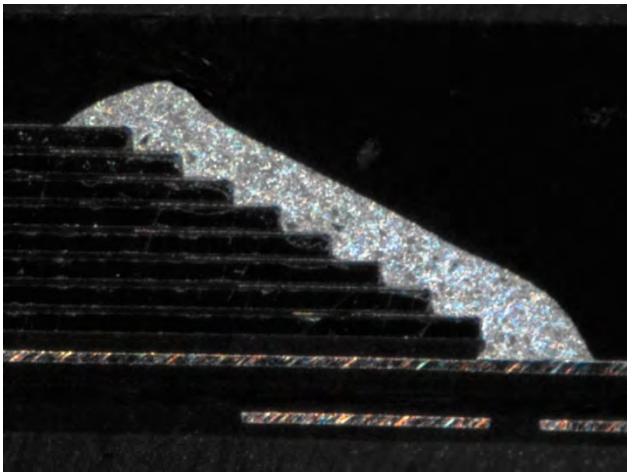


Figure 8: Cross-section of Conformal Conductor on Tier Stack of Flash Memory

WLP Methods for Cost Reduction & Cycle Time Reduction

Substantial packaging cost reductions are realized with a gang-based process due to the elimination of repeated die attach and pad-based wirebond processing steps, and the reduction or elimination of gold. Conformal vertical interconnect results in a 25% cost savings when an RDL layer is used to adapt existing die designs to the conformal interconnect design rules. When die are designed for use with conformal interconnect, negating the need for an RDL layer, cost savings of 75% can be achieved. Maintaining a wafer array format through many of the process steps, and maintaining a large array processing format after singulation occurs, makes all process steps amenable to the use of automated processing tools, spreading processing costs across large numbers of components, thus reducing the cost per component. In addition to reducing the total number of manufacturing operations and cycle time required to integrate a stack of die, this process requires less equipment, less manpower, and consequently, less manufacturing floor space. (Figure 9)

Ex. 1016, Page 802

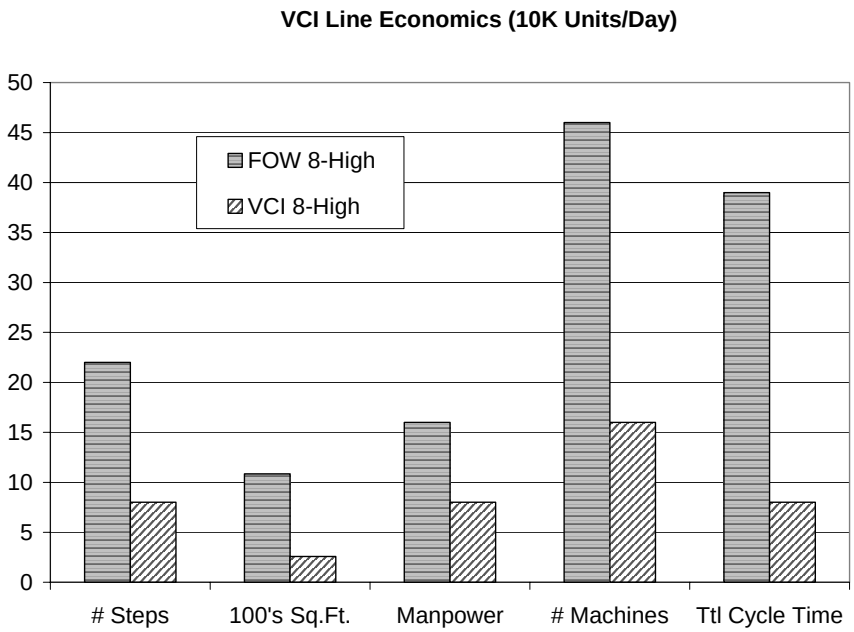


Figure 9: Manufacturing Facility Demands for 10K 8-high stacks/day

Edge Conductor Electrical Characterization

Electrical properties have been well studied and are quite adequate for the short interconnect distances between the die, exhibiting resistances in the tens of milliohms. For a die thickness of 50 microns, we can achieve a connection resistance less than 10 milliohms. Figure 10 summarizes measured resistance, inductance, and capacitance values.

VIP	R	L	C
Signal and power lines	0.171 ohm/mm	0.053 nH/mm	0.143 pF/mm
2-high DRAM stack (VIP height ~ 200um)	0.034 ohms	0.010 nH	0.028 pF
4-high DRAM stack (VIP height ~ 450um)	0.077 ohms	0.023 nH	0.064 pF

Figure 10: Edge Interconnect Characteristics

Summary

By utilizing a conductive filled liquid to form conformal vertical interconnects, significant savings of cycle time, costs, and size were achieved. Figure 11 summarizes and compares technical merits for conformal interconnect and traditional wire-bond approaches.

Ex. 1016, Page 803

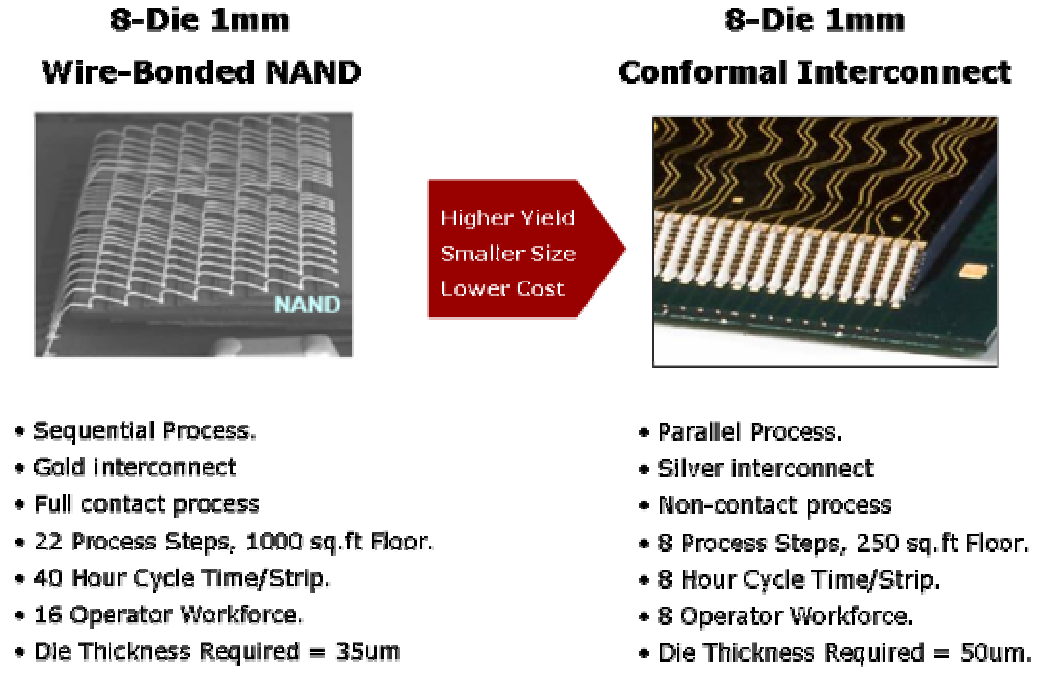


Figure 11: Summary of Technical Merits (10K 8-high stacks/day)

Acknowledgments

Simon McElrea, Jeff Leal, and others at VCI have contributed extensively to this research and development effort. We would also like acknowledge the support of our equipment and material suppliers Asymtek, DISCO Corporation, LORD Corporation, Resonetics, and Specialty Coating Systems.

References

1. J.C. Eloy, "Market Trends & Cost Analysis for 3DIC & TSV Integration", *2008 Pan Pacific Microelectronics Symposium, Jan 22-24, 2008, Hawaii*
2. P. Marcoux, "Improvements to Through Silicon Vias, or TSVs", *2008 Pan Pacific Microelectronics Symposium, Jan 22-24, 2008, Hawaii*
3. Lawrence D. Andrews, Terrence C. Caskey, Simon J. S. McElrea. *"3D Electrical Interconnection Using Extrusion Dispensed Conductive Adhesives"*, 32nd International Electronics Manufacturing Symposium, Oct.3-5, 2007, San Jose, CA
4. "SCS Parylene Specification and Properties", Cookson Electronics Equipment, Datasheet, Sept. 2002

Ex. 1016, Page 804

ECS Transactions, 18 (1) 777-785 (2009)
10.1149/1.3096536 © The Electrochemical Society

Next Generation Nickel-Based Bond Pads Enable Copper Wire Bonding

Bob Chylak[a], Jamin Ling[a], Horst Clauberg[a], and Tom Thieme[b]

[a] Kulicke and Soffa Ind. Inc.,1005 Virginia Drive, Fort Washington, PA 19034, USA
[b] Atotech Deutschland GmbH, Erasmusstraße 20, 10553 Berlin, Germany

Copper wire bonding has huge cost advantages over gold wire bonding. As a result, low pin count, heavy wire applications have already been converted to copper wire and many companies are in high volume production. Recently, with the price of gold skyrocketing, conversion of high pin count (>250 I/O), high performance applications to Cu has dramatically accelerated.

These high performance devices are increasingly relying on low-k materials under the bond pads. Unfortunately, the 33% greater hardness of Cu compared to gold places even greater stress on these inherently fragile materials. This can result in difficulties with pad damage and cratering of the underlying structures. Advancements have been made to copper wire bonders, tools, and wire that have resolved many of these issues and made fine pitch copper wire bonding feasible.

Ni-based bond pads have emerged to solve the pad damage problem. Nickel is about 50% harder than copper and four times harder than aluminum so that it provides greater protection against the higher stress resulting from Cu ball bonding, as well as damage during probe. This is especially beneficial for devices with low-k active circuitry under the bond pad. NiPd, NiPdAu, and/or NiAu have demonstrated their great robustness to receive the Cu wire bonding with a huge wire bonding window without any splash and with excellent reliability.

Introduction

Gold wire bonding has overwhelmingly predominated in the industry for more than twenty years because of its high reliability and performance. Other types of wire bonding, such as copper or aluminum, on the other hand, are more popular for heavy wire, low cost applications in consumer products or discrete semiconductor products. As gold prices have risen to historical highs, copper wire bonding has again gained attention for its advantages of low density, high stiffness, low cost, high electrical and thermal conductivities and excellent (intermetallic) reliability with aluminum pad metallization. Fundamental understanding and the infrastructure for fine pitch copper wire bonding has been steadily maturing.

Because of its oxidation sensitivity, a special copper bonding kit is necessary to apply an inert or reducing environment when forming a ball through melting prior bonding. Such copper kits have now matured and conversion of gold process to copper is readily achievable. Copper's high modulus relative to Au could also ease the requirements of

Ex. 1016, Page 805

long, skinny wire for ultra-fine wire applications. High electrical and thermal conductivities and excellent Cu-Al intermetallic reliability further enhance the appeal of copper wire bonding.

However, the high stiffness of copper also introduces difficulties during bonding. A higher bonding force is required to respond to its greater work hardening in order to deliver satisfactory bonding integrity. While applying ultrasonic energy during bonding, shear forces add to the already applied normal bonding force. Such shear stress can transmit through bond pad metallization to the brittle, easily fractured dielectrics underneath, leading to pad peeling and bond failure. High hardness of copper could also smear off the soft aluminum during bonding along the ultrasonic direction, which is commonly called 'Al splash'. Significant effort has been focused on optimizing the bonding parameters, such as ultrasonic energy and bonding force, to reduce such pad damage and Al splash with some level of successes. However, as the device technology progresses to finer bond pad pitch and thinner pad metallization, pad management deserves much more attention, not only for copper wire bonding but also for ultra-fine gold wire bonding.

Bond Pad Structure

Current bond pad structure migration

When the submicron technologies were introduced, multiple metal layer structures were developed to accommodate the demands for complex device functions. In order to structure such multiple metal layers, a planarization process including CMP (chemical-mechanical polishing), via and via filling (typically W or TiW) were adopted. The bond pad structures therefore include varieties of a sea of tungsten vias underneath the Al pads for better current carrying capability. 0.5% or higher Cu was added into the Al or Al-1%Si metallization for better electromigration resistance in ever shrinking geometries. The bond pad design rules at the time restricted any active structure underneath the bond pad to prevent any destructive action from wafer probe testing and/or wire bonding. However, a strong desire of form factor shrinkage led to aggressive BOAC (bond over active circuitry) development [1], pushing wire bonding and probing improvements in addition to robust bond pad structures to minimize concerns for device integrity.

For deep submicron technologies beyond 0.27μm, 0.13um feature sizes were demanded for further performance with faster, more powerful devices with reduced RC (resistance/capacitance) delays. Cu interconnect and low-k dielectrics were therefore introduced and have now matured. Lower dielectric constant dielectrics are typically porous composite materials that can only tolerate low stresses from bonding or probing. Therefore, Cu interconnect and low-k dielectrics are being used in the lower levels of devices where RC concerns are critical, but medium-k dielectrics such as FSG (fluorosilicate glass) with Al cap finish are used for the top one or two metal layers for power carrying capability and good probing/bonding robustness.

Ex. 1016, Page 806

Bond Pad Design for Cu wire bonding

More than a decade of experience with copper wire bonding has now led to a set of generic bond pad design rules. Ideally, a bond pad with thicker and/or higher yield strength bonding material is desired. As reported in the past [2,3], either alloying Al metal with different dopant types or dopant levels, or designing the Al pad structure differently can increase the composite pad strength and significantly reduce the risk of cratering or other damage to the underlying structures. However, changing the Al metallization or stacking metallization in wafer fab process is not trivial due to the complex process integration and stringent process reliability qualification requirements.

Therefore, a minimum Al pad thickness of 1µm is required irrespective of the alloy types (1%Si or 1%Si0.5%Cu). Typically when the Cu layer under the Al is thinner than 1µm, the Cu wire bonding process has to be well-optimized to achieve a suitable bonding window without pad damage. In contrast, an underlying Cu thickness of more than 1µm is very robust and generally delivers a good Cu wire bonding process. However, when Al pad thickness is thinner than 7000 Å or even approaching 5000Å, pad damage is almost always found during Cu wire bonding irrespective of the type of dielectric.

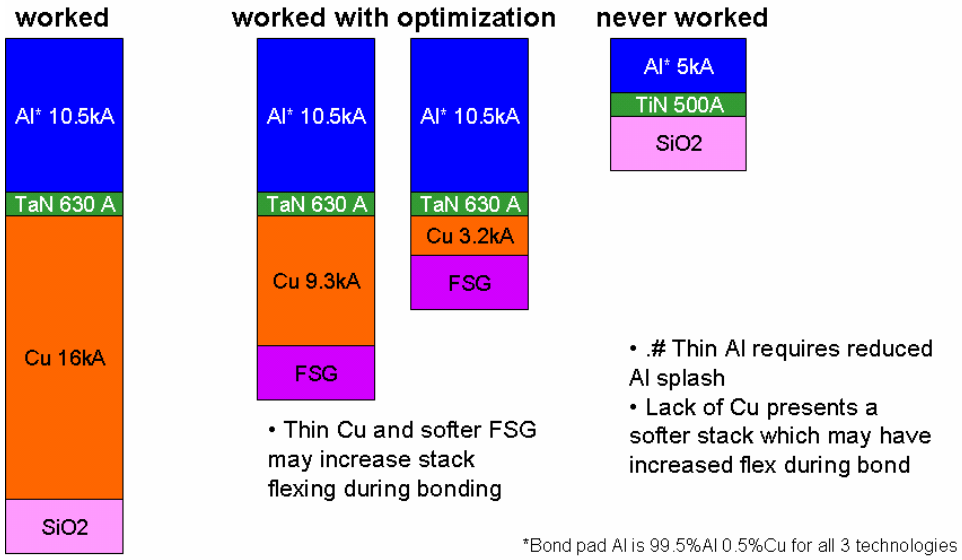


Figure 1. Pad Finish Stack Acceptable for Cu Wire Bonding

Dielectrics with a medium k, say above 3.2, such as FSG or TEOS SiO_2 typically have much stronger fracture resistance compared to low-k dielectrics. Bonding directly on top of low-k dielectric with either gold or copper wire bonding or even just probe testing such devices without pad damage has typically been found to be impossible [4, 5]. A tougher dielectric layer, either ILD (inter-layer dielectric) or passivation, has to be placed underneath the pad finish to enable robust wire bonding.

The use of a sea of via pad structure in conjunction with Cu wire bonding has been found to be detrimental. The tungsten vias in such W-SiO_2 composite structures serve as stress concentrators, causing cracks to propagate along the shortest distance between vias to release the pressure. Therefore, vias should not be placed under any high-stress areas of the bond pad. A peripheral via design for interlayer connection is much more desirable.

Ex. 1016, Page 807

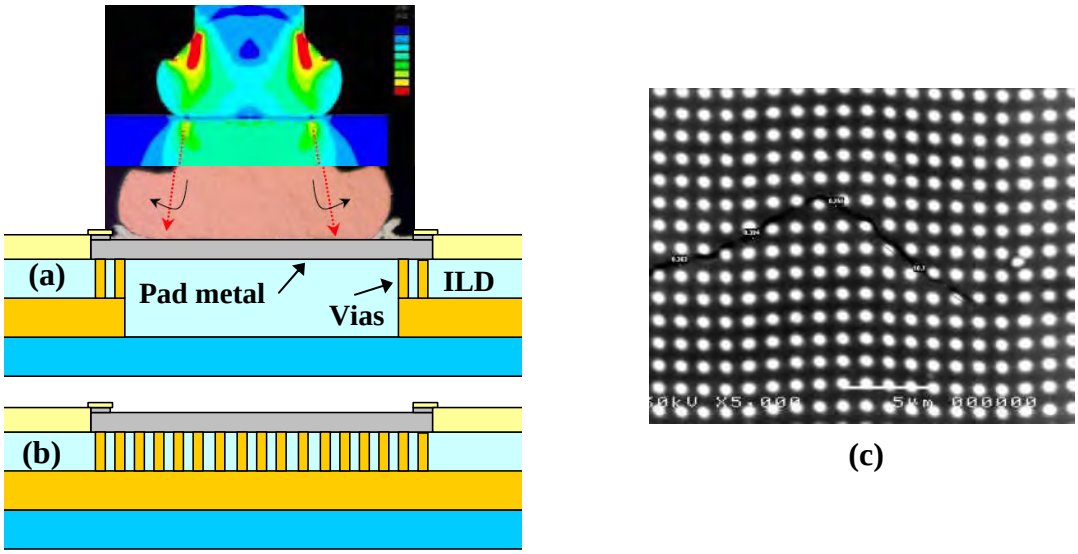


Figure 2: (a) workable bond pad via structure; (b) bond pad design with the sea of vias; (c) crack propagates between vias (top view, after pad metal stripped);

Bond Pad Solutions – Next Generation Ni-Based Bond Pad

The ideal next generation bond pad would address the needs of all of the emerging technologies, including copper wire bonding, ultra fine pitch gold wire bonding, ultra low-k dielectrics with rapidly shrinking BPP (bond pad pitch), as well as allowing probing and bonding on the same pad real estate.

Ni-based bond pad introduction

From a mechanical viewpoint for bonding and probing, bringing two materials with distinct material properties together obviously causes a higher amount of deformation in the softer material. As a result, aluminum splash is much more severe in Cu wire bonding than in gold wire bonding. Simultaneously, this deformation thins down the Al metallization underneath the bond, potentially inducing device damage and possibly reducing the bond reliability as the intermetallics grow.

Therefore, a harder pad metallization is desirable to protect the underlying structures. Nickel, with a hardness 50% and 70% higher than Cu and Au, respectively, is receiving increasing attention as the predominant alternative to aluminum metallization. It addresses all of the requirements of higher reliability, higher bonding load, protection of fragile structures, compatibility between probing and bonding, and compatibility with Au and Cu wire bonding [4,5,6].

Typically, 1µm to 3µm thick Ni is deposited on either Al or Cu base metallization. However, as a thin oxide rapidly forms on Ni, a thin noble layer of Au and/or Pd should be used on top of Ni for more robust manufacturability, bondability, and reliability. The combinations of NiAu, NiPdAu, NiPd structures have been matured and are used in many

other semiconductor or electronic applications. The layer thickness and stack structure combination may just need to be optimized to be suitable for the next generation bond pad structure of each specific application requirements.

Electroless plating bond pad

Electrolessly plated Ni-Pd and immersion plated Au has become an established technology in electronic component manufacturing. Recently it has become popular for wafer level, low cost flip chip under bump metal application because of it is low cost, maskless deposition and compatibility with Cu and Al substrates.

In this study, a test vehicle with Al pad finish was used. The bond pad pretreatment has to remove any residues or oxide layer from the aluminum alloy deposition process, followed by an alkaline etching to remove the top pad surface layer. Then, an immersion zincation process can provide the required surface protection as well as the necessary roughness characteristics for the subsequent electroless metal deposition. By replacing aluminum elementary with an immersion zinc layer, the surface is protected from forming any oxide layer.

Electroless nickel is then plated onto this zinc layer on top of the Al. The stress in the nickel layer is mainly influenced by the phosphorous co-deposit. Typical mid-phosphorous nickel layers can range from 7 to 9.5 %P and generate a tensile stress. Alternative electroless nickel types with a high P content of between 10% and 11% generate a compressive stress.

A rather new approach is an additional palladium intermediate layer between Nickel and gold or pure palladium as top metal layer. Acting as an additional protective barrier, the palladium significantly improves the bond pad stability, the corrosion resistance of the nickel and the stress impact of the metallization stack. The complete metal stack is more resistant to metal diffusion during thermal processing steps. Especially the stress impact can be minimized as the individual metal layer can be thinned down without risking nickel corrosion. Palladium not only allows minimizing the thickness of the nickel layer from 5µm to less then 3µm, but also eliminates the necessity of gold as the bondable pad surface finish.

Copper Wire Bonding on Nickel/Palladium/Gold Pads

Experimental

Nickel/palladium/gold layers were plated onto 1µm thick Al pads (Al-1%Si-0.5%Cu) of test wafers. An identical test wafer without any Ni/Pd/Au served as a reference. Bonding tests were performed using 0.8mil Heraeus Maxsoft wire on 50µm pitch pads at a 165ºC bond temperature. All bonding was performed using a K&S Maxµm Ultra wire bonder with a Microenvironment Copper Kit to provide forming gas (5%/95% H2/N2) cover gas during ball formation. The bonding tool was a K&S CuPRAplus capillary with 1.25 mil CD and 60º ICA. All devices were plasma cleaned in a March plasma cleaner for 5 min in Ar at 200 mtorr. Shear and pull measurements were performed using a Dage 4000 tester.

Ex. 1016, Page 809

Bonding Results

Screening study. In an initial screening experiment, devices with eight combinations of Ni/Pd/Au pads and an Al reference device (see Table I) were wire bonded with the same wire bonding parameter set. The bonding parameters were optimized on the Al reference and then applied without further changes to the Ni/Pd/Au pads. As Figure 3 shows, the shear/area response on all of the Ni/Pd/Au-plated devices was essentially the same with average shear per area between 12 and 14 g/mil^2 in each case. These shear/area values were about 6 g/mil^2 higher than those on the Al reference pads. The shear failure mode was through the Cu ball in all cases on the Ni/Pd/Au pads, but failure at the Al-Cu interface for the Al reference device. Top-of-loop pull strengths were between 9 and 9.5 g on all devices, including the Al reference. All wires broke at the neck of the balls.

Table I: Ni-Pd-Au Surface Finishes

Sample	Ni (µm)	Pd (µm)	Au (µm)
Al Ref	0	0	0
A	3	0.3	0.03
B	3	0.3	0
C	3	0.1	0.03
D	1	0.1	0
E	1	0.1	0.03
F	1	0.3	0.03
G	3	0.1	0
H	1	0.3	0

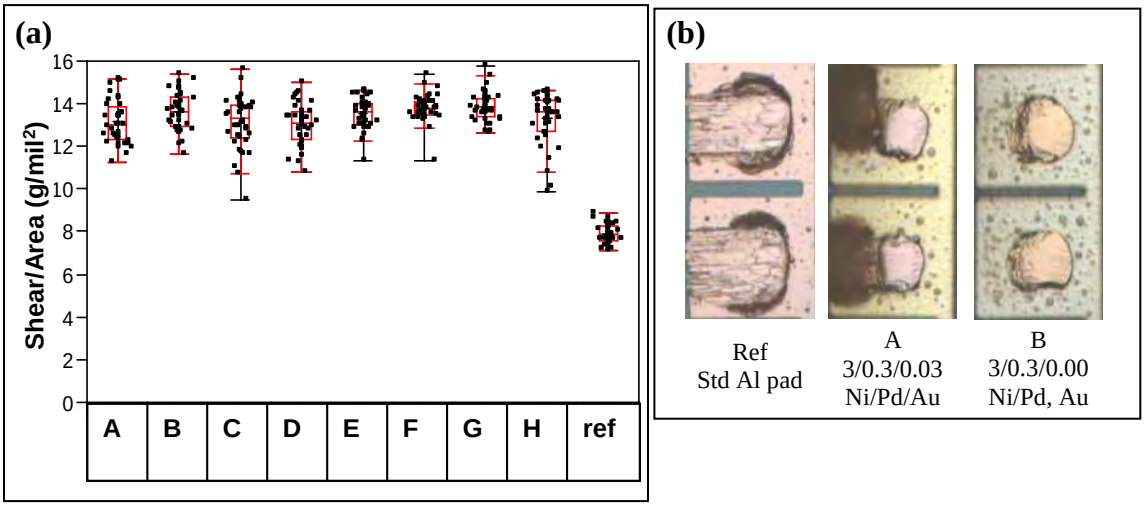


Figure 3: (a) Shear/Area for plating finishes in Table I (b) optical images of failure mode in shear test

The higher strength of the Ni allowed bonding without any of the pad splash (Figure 4(a)) that is normally seen in Cu wire bonding on Al pads. Clearly visible in Figure 4(b), such splash can add several microns to the effective ball diameter. The cross sections in Figure 5 of bonded devices demonstrate that there is minimal to no deformation of the Al layer under the Ni layer. For the thinner, 1µm Ni plating, a slight deformation of 0.2 µm

of the underlying Al pad is seen. The thicker, 3µm Ni plating completely protected the Al from any deformation. In contrast, the thickness of the Al under the Cu ball is often reduced by 80% or more on standard Al pads.

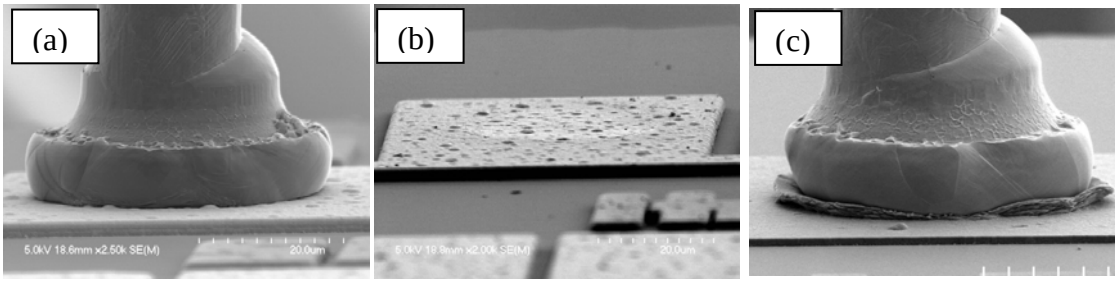


Figure 4: (a) Cu ball on Sample F (1um Ni) bonded at 85mA USG, (b) after acid removal of Cu ball from Sample F, (c) Cu ball on Al ref pad bonded at 80mA. Note lack of deformation of Ni/Pd/Au pad.

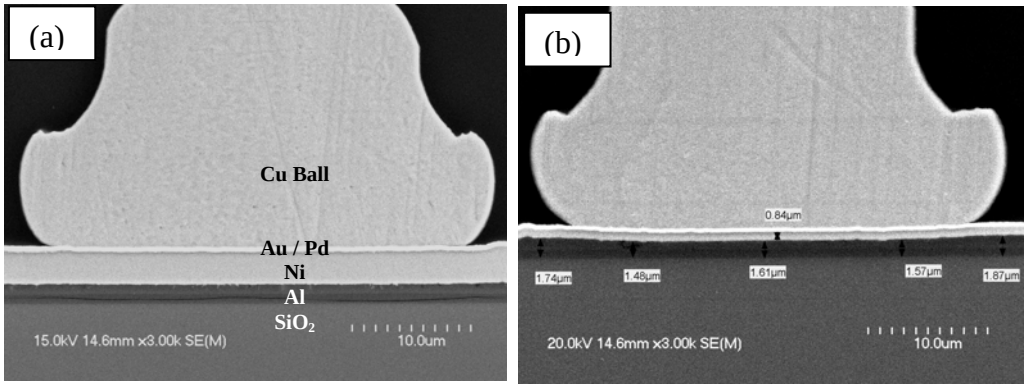


Figure 5:Cross-sections on Cu ball on (a) Sample A - 3um Ni, 0.3um Pd, 0.03um Au and (b) Sample F - 1um Ni, 0.3um Pd, 0.03um Au.

Ultrasonic Power Sensitivity Study

Figure 6 shows the results of an ultrasonic power sensitivity study in which samples A (3/0.3/0.03 µm Ni/Pd/Au), sample F (1/0.3/0.03 µm Ni/Pd/Au) and the Al reference device were compared. Ultrasonic energy was incremented in steps of 5 mA while keeping all other bonding parameters fixed. This type of experiments is typically used to establish the bonding window of a process (Table I).

The diameter of the copper balls responds very similarly to the ultrasonic energy on both Al and Ni-Pd-Au pads. However, as can be seen from the top line in Figure 6(a), the Al splash adds significantly to the effective ball diameter when bonding on Al pads. Since this splash limits the available space for the ball, it defines the upper limit of useful ultrasonic energies. For this typical 50µm pitch process with an upper spec limit of 40µm for the diameter, the ultrasonic energy is limited to 85mA when bonding on bare Al pads. In contrast, since there is no splash and no pad damage, ultrasonic energies as high as 110 and 115 mA can be used on pads with 1µm and 3µm Ni, respectively.

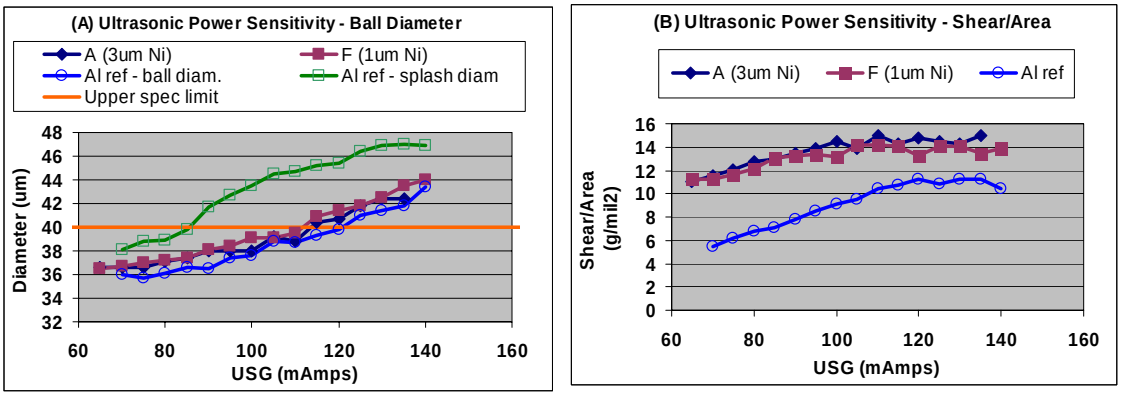


Figure 6: (a) Bonded ball diameter and Al splash diameter, (b) shear/area on sample A, F and Al reference

The lower spec limit of the bonding window is normally defined by the USG level that gives shear/area at the lower spec limit. This limit is usually set to 5.5 g/mil^2. On the Al pads, this lower ultrasonic limit was found to be 70mA. On the Ni-Pd-Au pads, shear/area was always far above the shear/area spec. Instead, the lower ultrasonic limit was set by the occurrence of NSOP (non-stick on pad). This limit was 70mA for both of the Ni-Pd-Au samples.

The ultrasonic energy range between the lower and upper limits is the bond window. As seen in Table I, the bond window for the nickel pads is approximately three times larger than that on the aluminum pads

Table II: Ultrasonic Bond Window (mA)

Sample	Min USG no NSOP	Min USG Shear > 5.5 g/mil^2	Max USG Dia. < 40µm	Bond Window
A (3µm Ni)	70		115	70 - 115
F (1µm Ni)	70		110	70 - 110
Al ref		70	120 for ball 85 for splash	70 - 85

Conclusions

The advantages of nickel-palladium-gold surface finishes to protect Al pads from the greater stresses of copper wire bonding have been demonstrated. Improvements are seen in elimination of pad splash, elimination of pad damage and a large increase in the bonding window. These low-cost plating finishes can be readily applied to existing wafer designs and provide an alternative to redesigning sensitive pads for Cu wire bonding. Future experiments will test the reliability of these Ni-Pd-Au bond pads when used with Cu wire bonding. Reliability of such pads in Au wire bonding has been shown to exceed that of Au-to-Al wire bonding [7].

Acknowledgments

Authors would like to thank Alan Slopey at K&S and Sophia Vietrich at Atotech for their dedicated work to support the Cu wire bonding and the electroless plating, respectively, throughout this study. Deep appreciation must be extended to K&S colleagues Matt Meyer, Shai Friedman, Jon Brunner and Frank Keller for their enthusiastic discussions.

References

[1] G Heinen, R.J. Stierman, D Edwards, 'Wire Bonds over Active Circuits', IEEE Electronic Component Technology Conference (ECTC) 1994
[2] TC Wei, AR Daud, 'Cratering on Thermosonic Copper Ball Wire Bonding', J Material Engineering and Performance, Vol.11(3), Jun 2002, pp.283
[3] L England, T Jiang, ,Reliability of Copper Wire Bonding to Al Metallization', IEEE Electronic Component Technology Conference (ECTC) 2007, pp.1604
[4] BT Ng, Ganesh VP and C Lee, 'Optimization of Gold Wire Bonding on Electroless Nickel Immersion Gold for High Temperature Applications', IEEE Electronic Packaging Technology Conference (EPTC) 2006
[5] WA Sasangka, AC Tan, 'High Temperature Performance Study of Gold wire bonding on a Palladium Bonding Pad', IEEE Electronic Packaging Technology Conference (EPTC) 2006
[6] P.Retchav, et al., 'Mechanical reliability of Au and Cu wire bonds to Al, Ni-Au and Ni-Pd-Au Capped Cu Bond Pad', Micro Electronic Reliability, 2006
[7] Asaf Hashmonai, Kulicke & Soffa experiment in progress (2008)

ECS Transactions, 18 (1) 787-792 (2009)
10.1149/1.3096537 © The Electrochemical Society

Consideration of Gold Thickness in ENIG UBM Application

M. Nozu , T. Shibata , K. Yoshikawa , S. Hashimoto

C. Uyemura & Co., Ltd., Hirakata, Osaka 573-0065, Japan

In the ENIG UBM formation technology on silicon wafer, even if the same gold thickness is measured by XRF, different solderability can be obtained from ENIG films prepared from different bath conditions. The deposit characteristics of immersion gold film are focused on, and the relation to the solderability is investigated. A uniform film having over 0.02um exhibits excellent solderability, but a film having a vertical growth deposition mechanism or excessive nickel corrosion does not. For the vertical growth deposition, the cause is considered to be the existence of a thin gold thickness area, and as for excessive nickel corrosion, a remaining oxide film and the generation of a P-rich layer can be considered to be the cause for solderability deterioration.

Introduction

Recently, the electroless plating process for UBM (Under Bump Metallization) formation technology on the silicon wafer has been gaining attention due to the advantages of low-cost and high productivity. Especially, demand for ENIG (Electroless Nickel / Immersion Gold) processes are increasing in the soldering application field, and therefore the ENIG process is focused on in this study. The demanded solderability in the soldering field includes solder wettability and solder joint reliability. As one of the effective factors to the solderability, the deposit characteristics of the gold film can be considerable. The deposition reaction of immersion gold is a displacement reaction between Ni-P metal and Au ions, so therefore the nickel dissolution process and the reduction / deposition process of gold are both effective to the gold deposition characteristic. On the other hand, in consideration of the soldering process, it can be considerable that the exposure or oxidation of the nickel is critical for solder wettability, and also the nickel corrosion, which can also be so-called black pad, is critical for solder joint reliability[1-5].
When controlling solderability is considered, the gold thickness measured by X-ray fluorescence analyzer (XRF) can easily be considered as a management item. However, different solderability can be obtained from ENIG films prepared from different bath conditions even if the same gold thickness is measured by XRF. According to the principle of measuring the thickness by XRF, the final value is an averaged film thickness whether it is uniform or not. This study focuses on this point, and the cause of different solderability is discussed in the viewpoint of gold film deposition characteristic. In addition, a gold sulfite type is chosen as the gold bath in consideration of environmental influence.

Experimental

Blanket wafers were used for the test wafers which were Al-Cu sputtered (Cu=0.5%) on silicon. The daisy chain type wafer having a φ 60μm open mouth covered by Si-N passivation were used for ball pull test.

Ex. 1016, Page 815

The plating process including pre-treatment were chosen from commercial products (EPITHAS process, Uyemura). Gold plating was utilized after 5μm of nickel plating. Various gold plating conditions were applied mainly by changing the working conditions of commercial product (EPITHAS TDS-20, Uyemura). DI water rinsing was applied between each process. After plating, the sample was dried by cool dry air.
Gold film thickness was measured by X-ray fluorescence analyzer (XRF, SFT-9550, SII, measuring diameter : 100μm). 100μm diameter lead free solder balls (M705, Senjyu) were used for solderability evaluation, and RMA type flux (529-D, Senjyu) was chosen. Solder wettability was evaluated by measuring the melt solder diameter after putting the solder ball on 3cm x 3cm blanket wafers. Solder joint reliability was evaluated both from the strength and the mode of ball pull testing (300μm/sec., Dage, series 4000). Ball pull test mode was evaluated as an average of points, which are represented as 10points for solder appearing, 0point for nickel appearing completely, and 5points for the intermediate results. Gold deposit characteristic was evaluated by FE-SEM observation with the top-down view of the deposition morphology and the cross section view of the film / interface conditions. The cross section was prepared by FIB cutting at the edge of the film, and FE-SEM observation was conducted by setting the sample vertically after once exposed to the room pressure. The depth profile of each element that consists the ENIG film was evaluated by Auger Electron Spectroscopy (AES, JAMP-7810, JEOL, Measurement area : 140μm x 140μm).

Results and discussion

Various gold thickness samples were prepared by changing the time, temperature, pH, and gold concentration widely outside from the optimum control range in order to represent various plating conditions. These conditions are selected as the conditions that various gold thickness are obtained from. Furthermore, in consideration of gold bath variety, “Bath-A” containing a different complexing agent was applied, and various gold thicknesses were obtained by changing the plating time. The feature of Bath-A is that a thick gold film can be obtained due to the high deposition rate and unsaturated film thickness. Solder wettability and ball pull test results obtained from these samples are shown in Fig.1 and Fig.2 (i)(ii), respectively. From Fig.1, three types are roughly classified from (a) to (c) according to the gold thickness in the view point of feature, Classified (a) is the type having excellent solder wettability over 0.02μm (shown in straight line, time and temperature), Classified (b) is the type having better solder wettability over 0.03μm (shown in dotted line, pH and gold concentration), Classified (c) is the type not having better solder wettability even over 0.05μm (shown in heavy line, Bath-A). Concerning ball pull test results, a significant difference was not obtained from the ball pull strength (Fig.2 (i)) in these evaluations, however, a slight difference was obtained from the ball pull mode, Classified (b) had an inferior result at thin film conditions, and Classified (c)

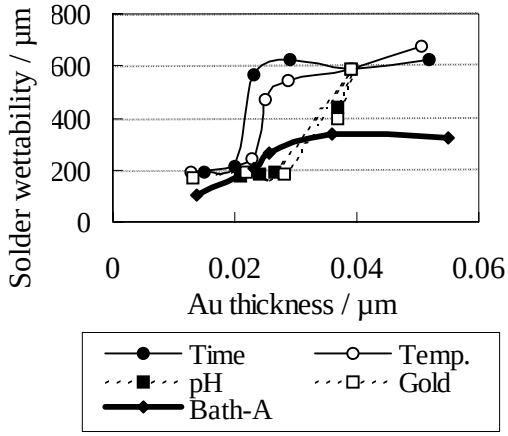


Figure 1. Au thickness v.s. solder wettability obtained from various bath conditions.

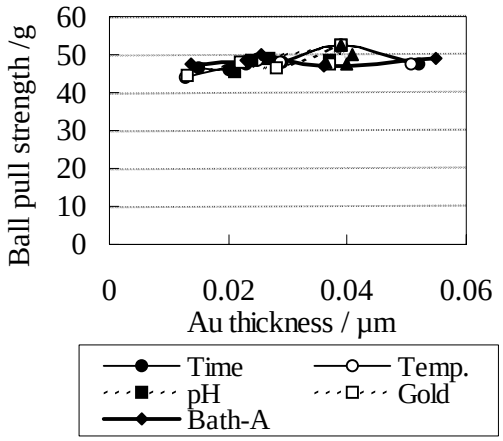


Figure 2(i) Au thickness v.s. ball pull strength obtained from various bath conditions

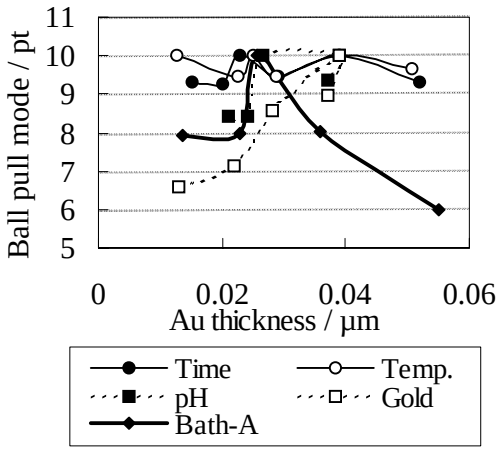


Figure 2(ii) Au thickness v.s. ball pull mode obtained from various bath conditions

had an inferior result at both thin and thick film conditions (Fig.2 (ii)).

In order to investigate in detail, three bath conditions were selected as a model case, standard condition of commercial product for Classified (a), low gold concentration (decreased from 2.0g/L to 0.75g/L) for Classified (b), and Bath-A for Classified (c), and various gold thicknesses were obtained by changing the plating time. The solder wettability result according to gold thickness is shown in Fig.3. Excellent solder wettability is obtained over 0.02μm from Classified (a), however, excellent wettability is not obtained from Classified (b) and (c) even in high thickness. Furthermore, the ball pull test result is shown in Fig.4. Ball pull strength does not show obvious tendency, but ball pull mode shows distinctive tendency. Classified (a) shows excellent pull mode, Classified (b) shows inferior tendency below 0.03μm, and Classified (c) showed inferior tendency over 0.04μm, not only in thin film conditions.

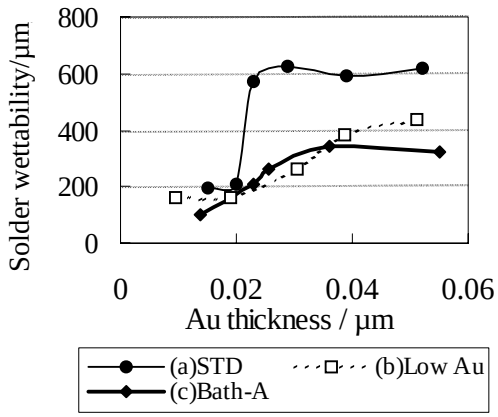


Figure 3. Au thickness v.s. solder wettability

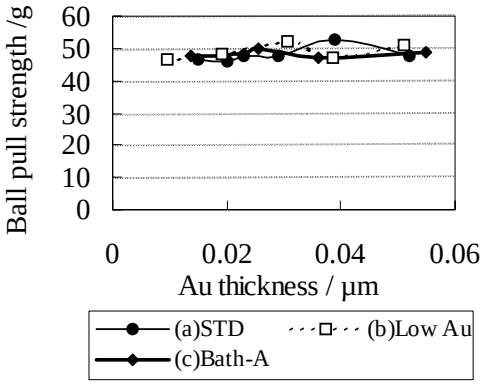


Figure 4(i). Au thickness v.s. ball pull strength

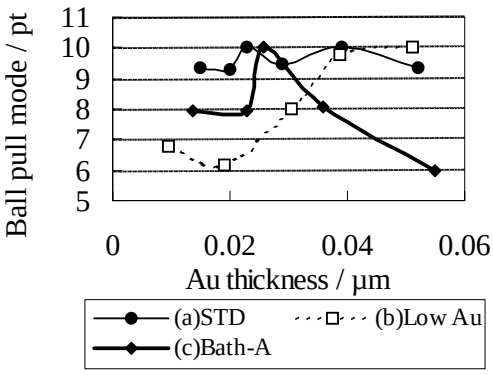


Figure 4(ii). Au thickness v.s. ball pull mode

Regarding these differences of solderability results, it can be considered that these difference come from the difference of gold deposit characteristic, concretely, gold film coverage, exposure of nickel, existence of oxide between Ni/Au, and generation of excessive P-rich layer, etc. Firstly, the relation between plating time and gold thickness is shown in Fig.5. Standard bath (Classified (a)) shows a saturation tendency, however, low gold concentration bath (Classified (b)) and Bath-A (Classified (c)) do not show a saturation tendency, and thickness goes up according to the plating time.

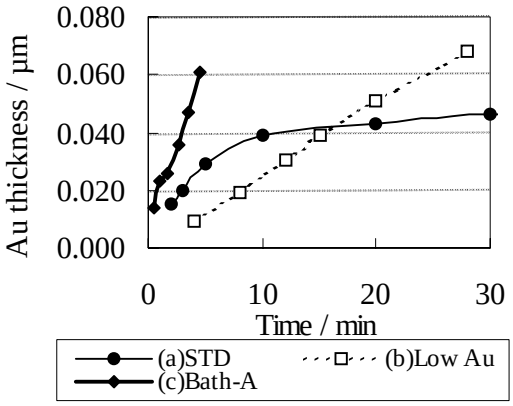


Figure 5. Plating time v.s. Au thickness

These results show that Classified (a) is due to excellent gold coverage with limited areas of nickel dissolution decreasing as gold thickness increases, however, Classified (b) and (c) are due to poor coverage of many points of nickel dissolution area. Secondly, in order to focus on gold deposit characteristics, FE-SEM observation of gold deposit morphology was conducted in two gold thickness conditions. The results from 0.02µm and 0.05µm are shown in Fig.6 and 7, respectively. Consequently, comparatively uniform deposition is observed from Classified (a) and (c), however, small particle deposition is observed from Classified (b). Furthermore, the cross section observation is shown in Fig.8 and 9. Classified (a) shows uniform deposition, however, as expected, Classified (b) shows rugged deposition. This result is in good correspondence to the expectation from Fig.5, that is, Classified (b) has a vertical growth gold deposition, and has a significant amount of nickel exposure and pits in the gold film, and consequently gold thickness does not show a saturation tendency. In the vertical growth deposition, the lowest film thickness is actually much lower even if the same film thickness is obtained by XRF, and poor solder wettability is easily considerable.

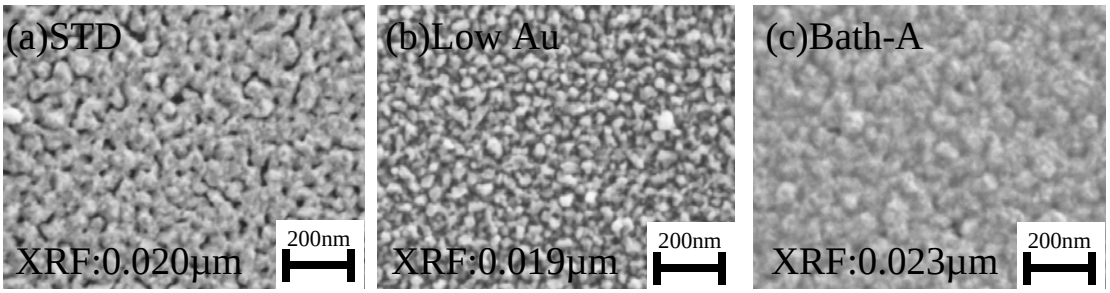


Figure 6. FE-SEM observation from each bath (Au thickness : 0.02µm)

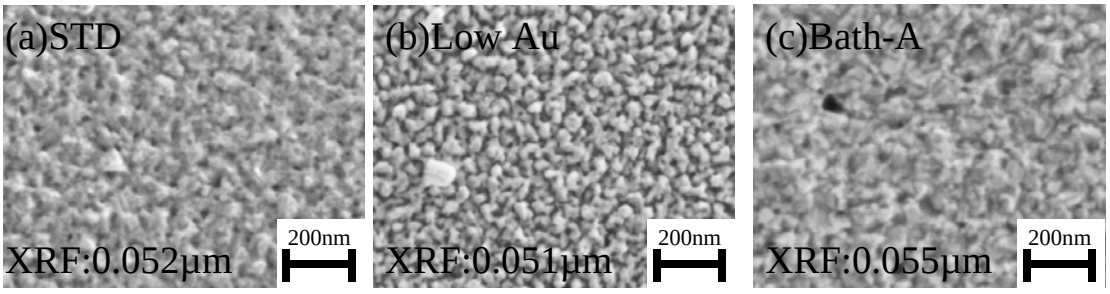


Figure 7. FE-SEM observation from each bath (Au thickness : 0.05µm)

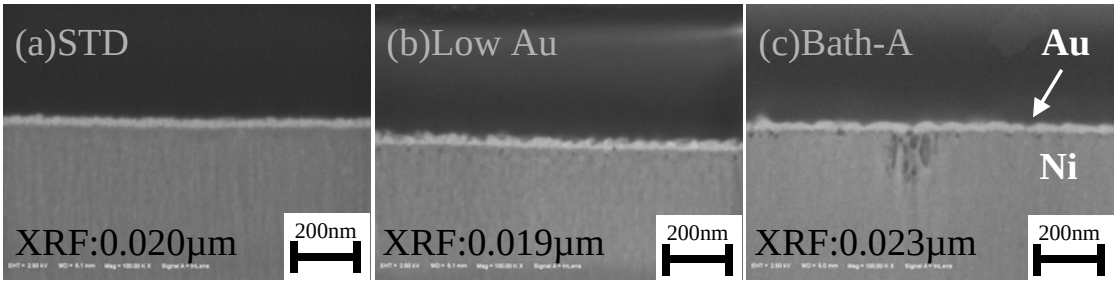


Figure 8. FE-SEM cross section observation of gold film obtained from each bath (Au thickness : 0.02μm)

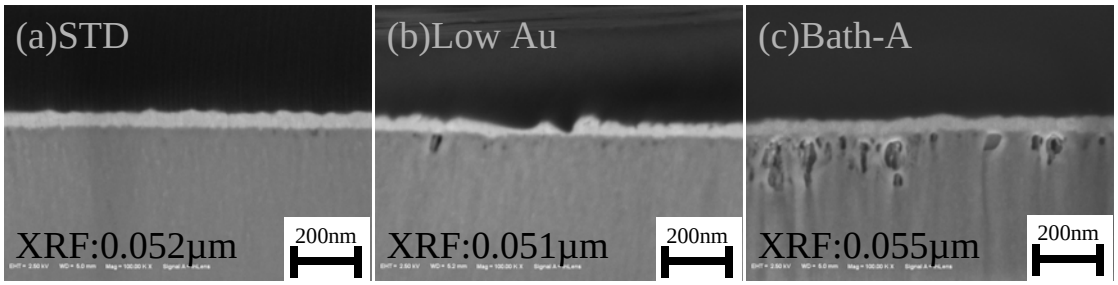


Figure 9. FE-SEM cross section observation of gold film obtained from each bath (Au thickness : 0.05μm)

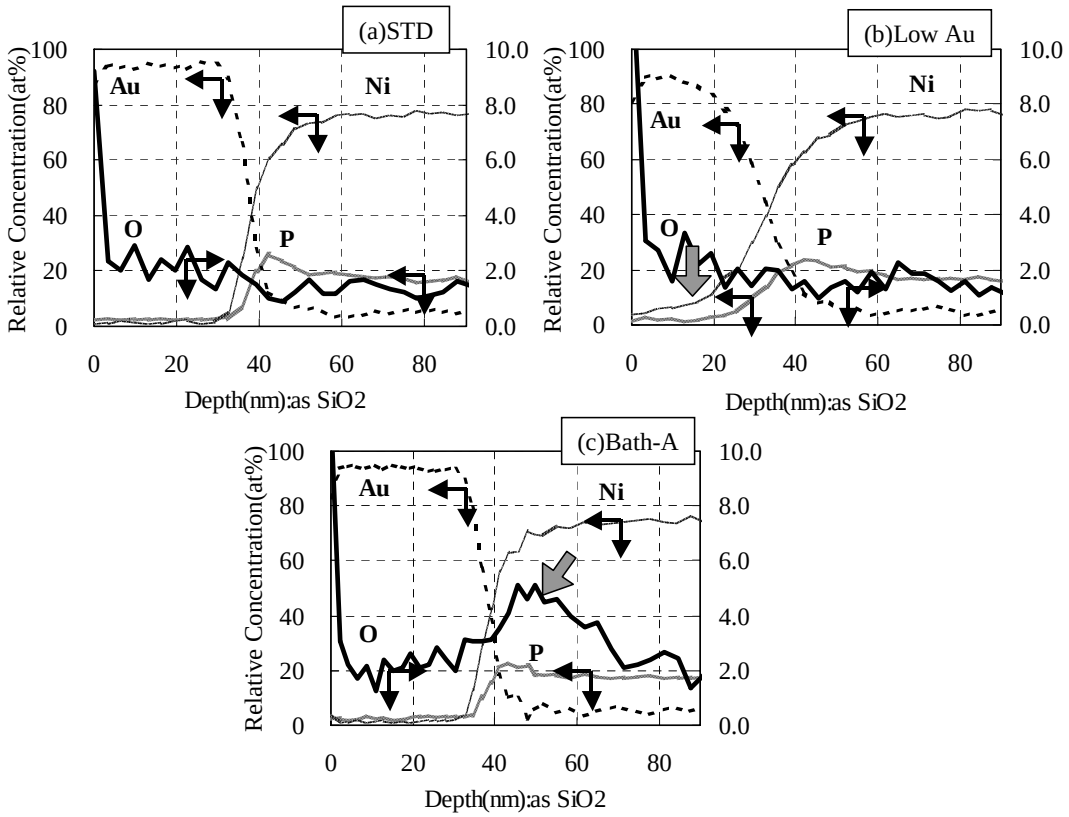


Figure 10. AES depth profile of the film obtained from each bath (Au thickness : 0.05μm)

On the other hand, excessive nickel corrosion is observed at the Ni/Au interface from Classified (c). Though it was not clear to be confirmed by this observation, it can be considered that a big pit of the gold film, as observed in Fig.7(c), exists in the portion where the nickel film has been greatly dissolved, and nickel is dissolved from this pit vertically, and consequently film thickness does not show a saturation tendency even though a uniform film is observed. The poor ball pull mode would be resulting from the generation of P-rich layer in the portion where the nickel film has been greatly dissolved. Furthermore, depth profile of the Ni/Au film is evaluated by AES. In comparison with Classified (a) having excellent solderability, the detection of the nickel was confirmed for Classified (b) at the early time stage when the gold film is detected. This result would show that as a result from vertical growth deposition, an area with a lower gold thickness exists, and consequently nickel is detected in the gold film. On the other hand, detection of oxygen is confirmed at the Ni/Au interface from Classified (c). This result would show that the gold deposition on the very surface occurred without enough dissolution of nickel oxide on the surface because the electrons are supplied from the vertical nickel dissolution. This oxide would make the solder wettability and solder joint reliability poor. From these results, it is suggested that different deposition characteristics are obtained from different plating conditions even with the same bath, and different solderability can also be obtained even if the same gold thickness is measured by XRF. It is confirmed that uniform gold deposition of 0.02μm shows excellent solderability, however, the gold deposition having vertical growth or excessive nickel corrosion does not show excellent solderability even in high gold thickness conditions.

Conclusion

In the ENIG UBM formation technology on silicon wafer, different solderability can be obtained from ENIG film prepared from different conditions even if the same gold thickness is measured by XRF. By focusing on the gold deposit characteristic, FE-SEM observation of gold surface morphology and cross section, and AES depth profile were conducted. If vertical growth deposition, excessive nickel corrosion, or the existence of oxide between Ni/Au are confirmed, poor solderability is obtained even from higher thicknesses. It is suggested that gold bath control or the selection of gold bath variation in the production line is critical, and also that 0.02μm of gold thickness is enough for solderability if uniform gold film is obtained.

References

1. E. Bradley and K. Banerji, in *Electronic Components and Technology Conference 1995*, p. 1028, Las Vegas, NV (1995).
2. G. Milad and R. Mayes, *Metal Finishing*, **96**, 42 (1998).
3. C. Key Chung, F. Mustapha, F. Hua, and R. Aspandiar, in *Electronics Packaging Technology Conference 2002*, p. 1, Singapore (2002).
4. N. Biunno, in *IPC Printed Circuits Expo '99*, p. S18-5-1, Long Beach, CA (1999).
5. R. A. Bulwith, M. Trosky, L.M. Picchione, and D. Hug, *Global SMT & Packaging Journal*, **2**, 9 (2002).

ECS Transactions, 18 (1) 793-801 (2009)
10.1149/1.3096538 © The Electrochemical Society

Dippable Paste: A Robust Assembly Material and Enabling Technology

H.G. Sy[a], R.R. Lathrop[b]

[a] Heraeus Materials Singapore Pte Ltd, Singapore
[b] Heraeus Contact Materials Division, W. Conshohocken, Pennsylvania 19428, USA

Despite the tackiness offered by ball attach fluxes, they have failed to prevent ball movement on Solder on Pad (SoP) finishes when the solder protrudes above the solder mask. Dippable solder paste provides the tackiness needed for these adverse situations as well as small solder particles that inhibit ball rolling due to mechanical or convective forces downstream. Also, other pad finishes like OSP has issues with oxidation with partial pad wetting as a common defect. The use of dippable paste has also demonstrated relief with these wetting related issues. The most current application of dippable pastes is in the assembly of stacked BGA's also known as Package on Package (PoP). Although water clean formulations are typical in BGA assembly, PoP assembly typically utilizes no clean formulations. Although tacky flux has been successful in many PoP applications, issues such as coplanarity and package tolerances that require increasing the sphere size during assembly have found benefit from dippable solder paste.

BGA Ball Attach Defect

Many of the commercial tacky fluxes available today coupled with highly accurate ball placement machines have been able to further improve the first pass yields since its beginning with manual ball attach. Today the typical first pass yield for FCBGA packages with body sizes of 15mm to 45mm and ball counts ranging from 200 to over 2000 is 99%.

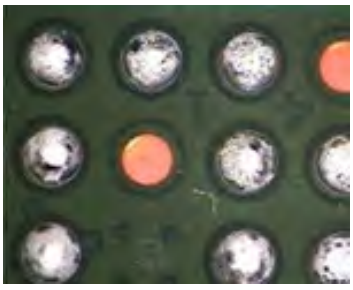


Figure 1 Image of Missing Ball



Figure 2 Image of Joined Balls



Figure 3 Image of Bridging Ball

To this day, the exact root cause of missing balls remains unsolved using the tacky flux only process. It is thought that a combination of factors such as variations in the pad finish condition, flux activator content and outgassing, amount of flux deposition, solder ball oxidation, transport vibration and reflow conditions contribute to the typically less than 5,000ppm of the missing balls defect rate. A systematic study was done by Chiu and Lee on typical ball attach defects such as missing, misalignment, and bridging, and it

Ex. 1016, Page 821

was concluded that the use of solder paste produces a higher yield (1). To this day however, the application of solder paste for BGA ball attach is limited. A typical application of solder paste printing is to attach high lead balls on ceramic substrates where eutectic SnPb (63/37) alloy paste is used.

With the tacky flux process, joined balls and solder ball bridging are typically caused by bridging flux, flux flooding or decreasing flux viscosity between solder balls combined with ball movements. Ball movements can be caused by transport vibrations from the ball placement machine to the reflow oven as well as by hand or automatic conveyor systems. Furthermore, as the flux slumps when it goes through the reflow oven, small vibrations of the reflow conveyor belt and uneven airflow in convection ovens can cause the balls to move and wick towards one another. The warpage of the package will also magnify this effect because it has been found that larger FC package sizes are found to be more prone to solder ball movement than smaller ones as demonstrated by lower first pass yields with these packages.

Ceramic BGA substrates have flat pads that are not mask defined. The edge of the mask on the solder pad in PBGA packages helps to limit ball movement. For this mechanical reason CBGA packages are also quite prone to ball movement.

To prevent ball movement, many flux suppliers have improved the tackiness of the fluxes while maintaining or improving its activity. However, until today, no tacky flux formulation is commercially available that prevents all ball movement. Typically, tacky fluxes would begin to slump at increasing temperatures whether the method of application is by stencil printing, silk screen printing, or pin transfer.

In some instances for packages with Solder On Pad BGA pad finish (SOPBGA), the solder height is significantly above the surrounding solder mask. These pad locations are quite prone to solder ball rolling which results to either joined ball or solder ball bridging.

Recently, a breakthrough water soluble solder paste with very fine metal powder applied by pin transfer application methods, also known as a “dippable paste” was developed. The results of replacing tacky flux with this unique material was that no missing, bridging, or joined balls have been reported (2).

With the dippable paste, the very small particles, usually 5-25um, act as “wheel chocks” that prevent ball movement through the reflow oven. Additionally, it is thought, the fine solder particles allow better contact with the base metal during the initial stages of reflow as they melt before the ball.

Although the tackiness is lower than some of the tackiest fluxes available it has shown remarkable capability to prevent ball movement. In a pilot study of 100 units with comparing flux to dippable paste on flat Ceramic FCBGA, the dippable paste showed 100% yield while the flux only, 98% (2).

Further validation on another 100 unit sample size conducted for this study with dippable paste on Ceramic FCBGA, 50mm package size, 1.0mm pitch, 0.6mm pad size, 0.64mm solder ball diameter, and 1156 ball count showed 100% yield. In this validation run there were no process issues encountered such as inconsistent deposit, ball movement, first pass defects, or paste pin bridging. The small sample size for validation work was limited due to resource constraints. This was, however, deemed sufficient data as ball movement was altogether eliminated.

From these trials it is suggested that dippable paste will even work for non-solder mask defined (NSMD) pads with 0.5mm or for even finer pitches such as 0.4mm and 0.3mm. It has been found that NSMD pads have greater reliability than solder mask

Ex. 1016, Page 822

defined pads (SMD) due to the anchoring effect of the solder around the copper pad. NSMD application is especially useful in CSP's for mobile phone applications.

Surface Finish Issues

There are various BGA ball attach pad finishes today and each one has its own advantages and disadvantages. NiAu finish has excellent solderability but is known for the "black pad" defect or "balls-off" failure. Immersion Sn is relatively cheaper and is uniformly flat but easily oxidizes and has a short shelf life. Solder on pad has excellent solderability but a little costly and prone to solder height variations.

Today, CuOSP pads are favored over NiAu pads for CSP's due to evidences of better drop test reliability performance, this is true with SAC305, SAC405, SAC125Ni, and polymer core solder balls (Sn3.5Ag) (3, 4). So it is foreseen that CuOSP pad will continue to be the surface finish pad of choice not just for reliability improvement but also for cost savings and simpler more environmentally friendly manufacturing process (5). There are drawbacks to CuOSP in that it is sensitive to thermal, humidity, and physical effects resulting to poor wetting (6). These drawbacks are the main reasons that for many years, assembly packaging houses have avoided CuOSP pad finishes.

The problem most assemblers face today is that commercially available ball attach fluxes are not able to deliver full wetting on CuOSP as OSP easily degrades as it goes through the upstream thermal processes such as die attach, wire bonding, molding, chip attach, baking, reflow, and plasma. Figure 4 shows partial wetting using commercial flux, detected at final visual inspection. In some cases, partial wetting may only be detected by zone shear as in Figure 5. Zone shear is a qualitative destructive method used to inspect for partial wetting. All of the balls are sheared by sweeping them simultaneously at 5mm/sec speed and the shear modes are inspected under a stereoscope. Typical shear mode failures are bulk solder failure mode, pad lift, IMC failures, and exposed copper due to partial wetting.

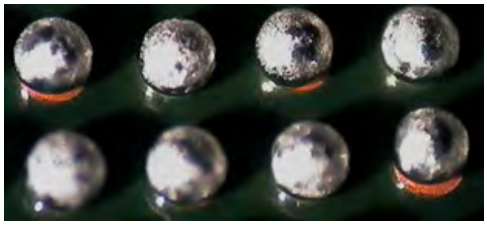


Figure 4 Partial Pad Wetting on CuOSP

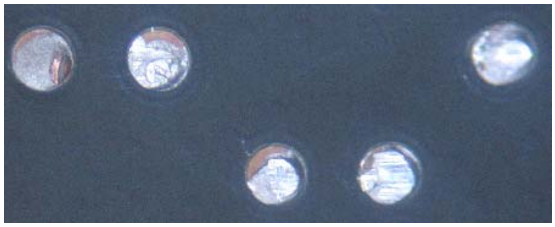


Figure 5 Partial Wetting Failure Mode after Zone Shear

The problem of partial wetting can be attributed to the fast oxidation of copper pads due to either poor OSP chemistry, non-uniform OSP coating, non-suitable flux chemistry or process issues like poor flux coverage on pad.

To remove Cu oxidation or to clean up CuOSP degradation, pre-cleaning of substrates is sometimes employed. Pre-cleaning involves the application of flux onto the substrates then washed away to remove the oxidation on the Cu pads due to OSP that has deteriorated from various thermal processes. The problem with pre-cleaning is that it becomes an extra process step increasing manufacturing cost. Depending on the cost considerations, some assemblers do not pre-clean because the rework process may cost more than the cost of the yield loss. Additional units are made to make-up for the lost volume in these situations.

Ex. 1016, Page 823

Dippable Paste Development

There are two main reasons why dippable paste inhibits ball movement, much higher total tack energy and small spherical solder powder. Tack energy is the total energy under the force–displacement curve in a tack test. As can be seen in the chart of Total tack energy in Figure 6, tacky fluxes have moderately improved over time. Tacky flux A is a formulation from several years ago and tacky flux B is the current state-of-the-art. The type 5 powder dippable flux provided another step up in tack energy form tacky flux and as the powder size was decreased (type 6) a significant increase was measured. This is also a common phenomenon in printable solder pastes and is due to surface tension effects for increased surface area of smaller diameter solder powder.

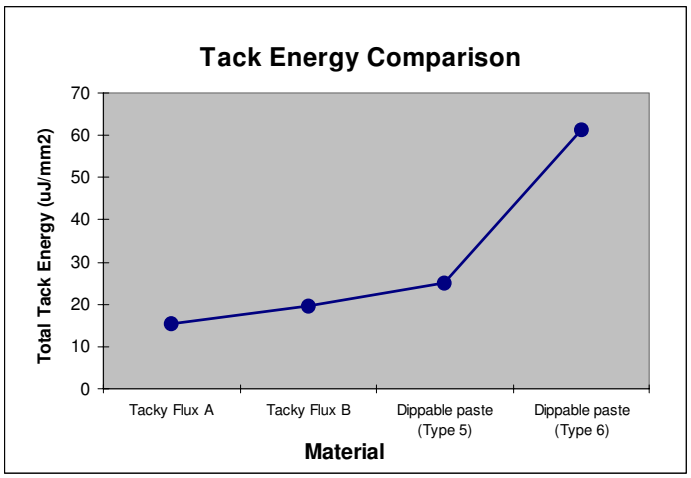


Figure 6 Total Tack Energy

The second benefit is that dippable paste has small highly spherical solder powder particles. These provide an mechanical stop or "chuck" to keep the ball from rolling. If they are too large they may roll with the solder sphere that is why type 5 and 6 powders are typically used. The quality of the powder is also critical and difficult to maintaining in conventional powder manufacturing methods with type 5 and smaller powder sizes. The Welco process is a unique patented process that yields extremely spherical powder as can be seen in Figure 7.

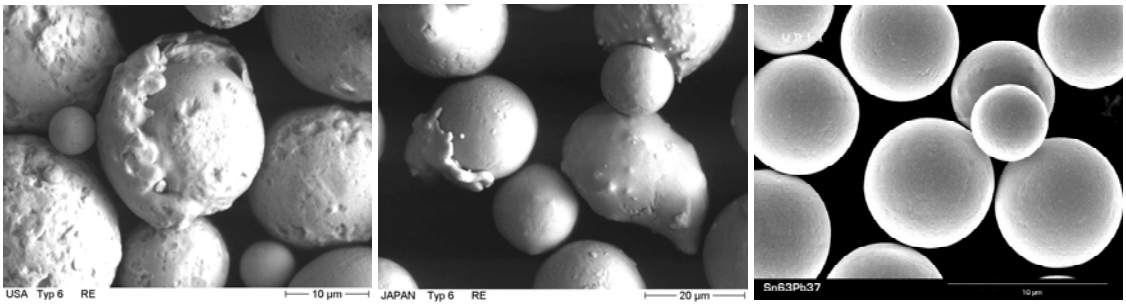


Figure 7 Type 6 Powder Quality Comparisons

Although dippable solder paste initially began with the ball dipping process, it quickly migrated to the pin transfer. For solder paste development "pin transferability"

was a completely new property to test. Although initially accomplished at equipment manufacturer and the end users, in-house development was the long term goal.

Various paste rheologies, metal loading and powder sizes were tested for pin transferability. A simple fixture loaded with various pin diameters and tip shapes was developed with spring loaded pins. Pins were loaded with dippable paste in a bottom justified adjustable depth reservoir then transcribed onto float glass. A confocal measuring system was used to measure the transferred deposits shape, diameter, height and volume. A run of 50 dips was done for each material under test with the 1st, 2nd, 10th, 20th and 50th dips measured. Between each dip and stamp cycle the reservoir was doctor bladed to replenish the reservoir surface with fresh material. Probe tips were photographed for wetting behavior after the 50th dip.

The method described was useful in determining the effects of metal loading on pin transfer performance. As can be seen in the transfer height charts in Figure 8, the flat tipped pin, typical of pin transfer tooling, took longer to stabilize the transferred height with the higher solids material. By the 50th stamp, however, the heights of the two different metal load materials were nearly identical. Too high a metal loading resulted in poor pin wetting and too low a metal load resulted in insufficient powder density in the transcribed deposit.

If only the 50th dip data is charted, it becomes obvious that the pin shape and size becomes significant (7). When the 525µ flat and round pins are compared, the data indicates that the flat tipped pin transfers much more volume with a much larger diameter as can be seen in the charts in Figure 9. For all of the round tipped pins the transfer is a clear trend of smaller deposits from smaller pins as anticipated.

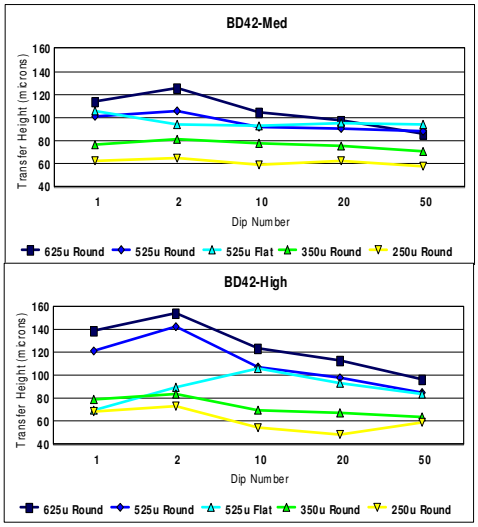


Figure 8 Metal Loading Testing

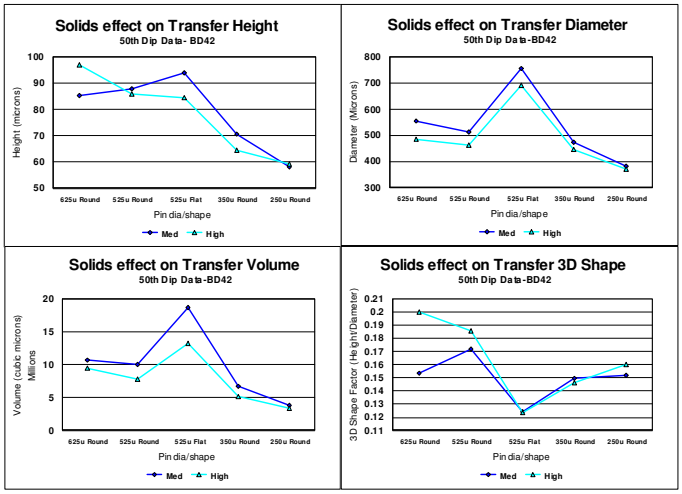


Figure 9 Pin Size and Shape Effects

Other factors considered in the development of water washable dippable solder pastes were cleanability, dip reservoir pot life and wetting. Cleanability was studied on a variety of different mask surfaces with dark, high surface area masks providing the most challenge. With trapped flux under low standoff components was not a concern. This permitted the inclusion of thixatropic additives to ensure an 8 hour pot life. Wetting was ensured in that the base flux chemistry for these materials had proven success on heavily oxidized copper OSP applications

Fine Pitch and Pin Transfer

Ex. 1016, Page 825

The dippable paste was originally developed to overcome ball movement and to be able to relieve OSP related issues for BGA ball attach applications in a ball dip method. In this method balls are dipped directly into a reservoir of dippable solder paste, and then placed on the substrate. Because of its very consistent rheological properties, it was also found that it prevents bridging between pins in a pin transfer method. In this method a fixed array of pins are dipped into a reservoir of dippable paste, and then contacted to the substrate. Balls are then placed on these "transcribed" deposits of dippable paste. The problem with tacky fluxes in the pin transfer method is that they tend to wick up the pins and bridge the pins over time. This results in manufacturing downtime as there is always the need to wipe off the bridged flux on the pins.

The workability of dippable paste was tested with an SCSP of 13 X 13mm body size, 289 I/O with a CuOSP finish pad, 0.5mm ball pad pitch, and a 0.25 mm solder mask opening. Figure 10 shows the formation of the dippable paste deposits on the 0.25mm round pins. Notice the defined rounded shape of the dippable paste on the pin tip. However, there was an observed variation with the transcription on pad as shown in Figure 11. This could be attributed to less than optimal pin transfer parameters such as dip depth and transfer pressure.

The consistency of deposit formation on pin tips of the dippable paste can also be observed in Figures 12a and 12b on a 0.5mm pin size with 1.0mm pitch.

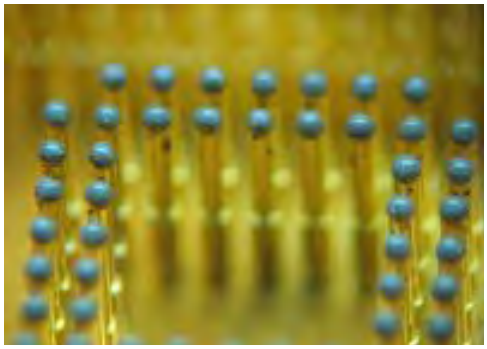


Figure 10 Image of 0.5mm Pitch with 0.25



Figure 11 Image of Dippable Paste

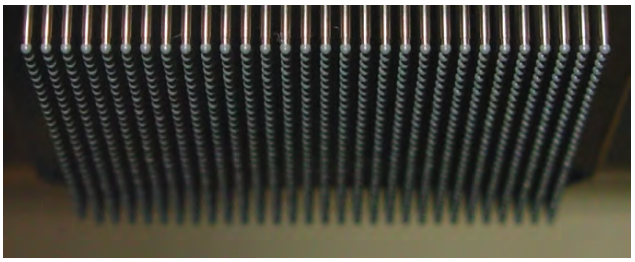


Figure 12a Image of 1mm Pitch Pins with Dippable Paste

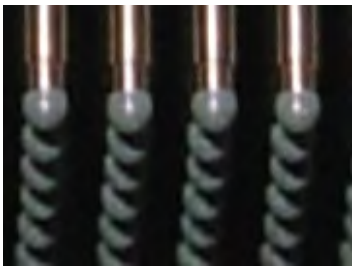


Figure 12b Paste Formation on Pin Tips

This typical formation of dippable paste on pin tips makes it suitable for fine pitch applications where flux only application is not suitable. This is foreseen as the enabling technology for finer pitch CSP/BGA ball attach process. A recent development study on 0.3mm pitch CSP/BGA showed a 91% first pass yield on 150 units sample size with flux printing process (8). Logically, flux flooding and flux bridging will be avoided by using dippable paste with pin transfer. Further experimentation with 0.4mm and 0.3mm is planned in the near future to validate this prediction.

Ex. 1016, Page 826

Stacked BGA

Stacked BGA, also referred to as Package on Package (PoP) is becoming more popular as a 3D package option due to its simplicity. The assembly of the two halves of this package design is typically done in level 2 assembly contrary to most package assembly. The lower package is placed in standard solder paste along with all of the other SMT components.

The upper PoP device assembly onto the lower device can be effectively accomplished with a dip application of tacky flux onto the upper PoP BGA before placing on the lower device. There are situations, however, where replacing tacky flux with a dippable solder paste (Figure 13) has been shown to be more effective. Figure 14 shows a typical BGA based PoP assembly. PoP applications where there is a need to increase the upper PoP devices ball size to increase or enable an adequate gap between the lower device over-mold cap and the lower surface of the upper PoP device, the solder particles (typically type 5 or 6) simply add metal to these balls during reflow. Other cases for using a dippable paste over tacky flux include compensating for device coplanarity and more recently ball to ball joining.

As with virtually all fine powder solder paste applications, nitrogen reflow is recommended to prevent solder balling. This holds true for both no clean and water clean formulations due to the exponentially high surface areas of type 5 and smaller powders. Figures 15 & 16 show a low residue dippable solder paste PoP application in both air and inerted reflow respectively (9).

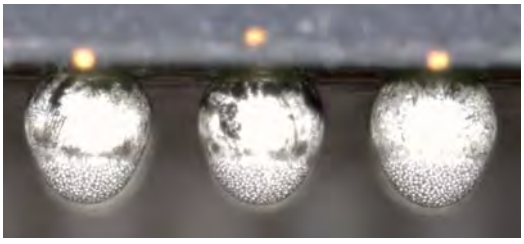


Figure 13 Dippable Paste on Upper PoP Device



Figure14 Typical Stacked BGA PoP Assembly

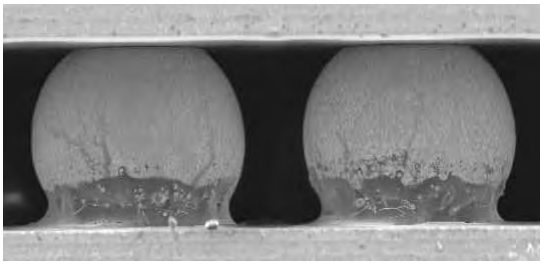


Figure 15 Air Reflow (upper Device)

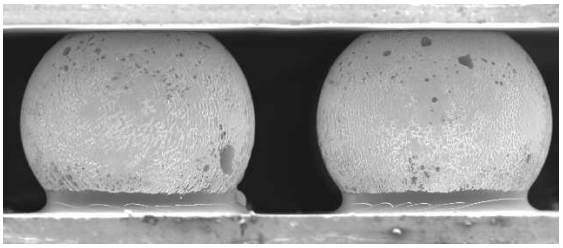


Figure16 Nitrogen Reflow (upper Device)

Material usage was measured by first weighing a dipped upper device then cleaning, drying and re-weighing. The microgram per ball usage is very similar for two different flux systems at the same metal loading for several different dip depths as plotted in Figure 17.

As with ball attach using dippable solder paste, the complete miscibility of the solder powder into the ball occurs during reflow as evidenced by the cross-section in Figure 18. Also the complete lower device pad wetting and minimal IMC thickness are visible in this photograph.

Ex. 1016, Page 827

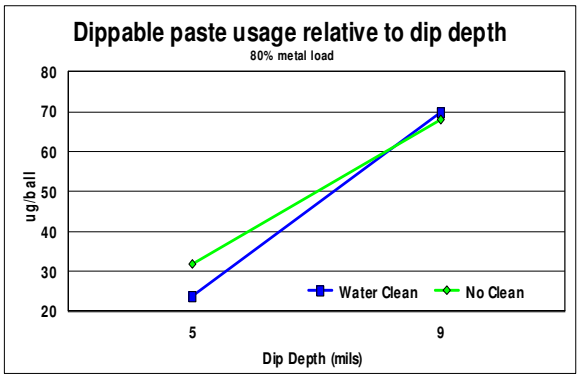


Figure17 Material Usage

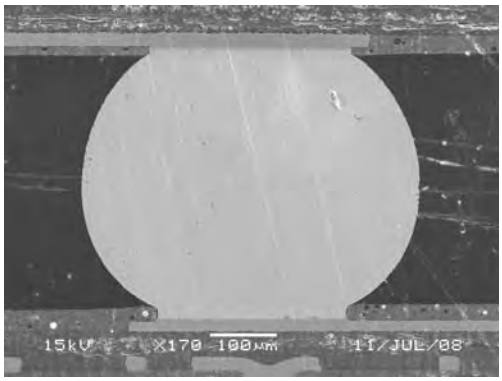


Figure 18 PoP Assembly X-Section

Conclusions

1. The dippable paste with pin transfer solved BGA ball attach inherent problems with flux only process such as missing balls, bridging balls, and joined balls regardless of pad design, variability, and substrate pad finish.
2. It is foreseen that dippable paste will enable finer pitch technology such as 0.3mm and below for CSP/BGA and allow designers to use NSMD ball pad design which is much needed for drop reliability improvement.
3. In developing the dippable paste, both an upper and lower metal load limit was established by simulating the pin transfer process. Too high a metal load prevented proper pin wetting and too low yielded powder density issues in the transcribed deposit.
4. For PoP assembly both water and no clean formulations can be used when the upper ball size needs to be increased or to compensate for device coplanarity.
5. All fine powder solder paste applications require inerted reflow to prevent solder balling.

Acknowledgements

The authors would like to thank Steve Fritzinger of Heraeus CMD, and Miguel Jimarez, ChangGyun Ryu, Jackson Hsu, Lito Mendoza of Amkor Technology and Daniel Tong of Zen Voce Company for their support.

References

1. Chiu, C. S. and N. C. Lee, *''Options and Concerns of BGA solder bumping''*, in Proc. of The Third International Symposium of Electronic Packaging Technology, Beijing, China. August 1998.
2. Sy, Hansen, Jackson Hsu, and Miguel Jimarez. *"New Robust Process Improvement for BGA Solder Ball Attach First Pass Yield."* Electronics Packaging Technology Conference Proceedings 2007. Dec 2007.
3. Syed, Ahmer, TaeSeong Kim, and SeWoong Cha. "Alternate Solder Balls For Improving Drop/Shock Reliability" SMTA International**.** Orlando, Florida. October 2007.

4. Chong, Y.R. Desmond., F.X. Che, John H.L. Pang, Kellin Ng, Jane Y.N. Tan and Patrick T.H. Low. *"Drop impact reliability testing for lead-free and lead-based soldered IC packages."* Microelectronics and Reliability, Volume 46, Issue 7, July 2006, Pages 1160-1171.
5. Chang, D.; Bai, F.; Wang, Y.P.; Hsiao, C.S. *"The Study of OSP as Reliable Surface Finishes of BGA Substrate"* Electronics Packaging Technology Conference, 2004. EPTC 2004. Proceedings of 6th Volume , Issue , 8-10 Dec. 2004 Page(s): 149 – 153.
6. Yuan Li. *"An Experimental Study on Organic Solderability Preservative"* Electronics Manufacturing Technology Symposium, 1997. Twenty-First IEEE/CPMT International Volume , Issue , 13-15 Oct 1997 Page(s):56 – 61
7. Lathrop, R.R., *"Pin Transferability Testing of Dippable Solder Paste Formulations"*, Proceedings of SMTA International Conference on Soldering and Reliability, Toronto, March 2008
8. Katahira, Takayoshi, Joan Scanlan, JongWook Park, and KwangSeok Oh. *"0.3mm Pitch CSP/BGA Development for Mobile Terminals."* 40th International Symposium on Microelectronics Proceedings. San Jose, California. November 11-15, 2007. pg. 393 – 401.
9. Lathrop, R.R., *"Semiconductor Packaging Solutions Utilizing Fine Powder Solder Pastes",* Proceedings of International Wafer Level Packaging Conference, San Jose, CA, October. 2008

ECS Transactions, 18 (1) 803-812 (2009)
10.1149/1.3096539 © The Electrochemical Society

The Destructive Wire Bond Pull Test Methodology for the Ultra Low Loop Application on Thinner Die Package

Bob Chylak[a], Oranna Yauw[b], O Dal Kwon[c], Kuan Fang Lee[d]

[a] VP of Global Process Engineering, Kulicke & Soffa, 1005 Virginia Drive, Ft. Washington, PA 19034, USA
[b] Manager of Process Development Engineering, Kulicke & Soffa, 6 Serangoon North Ave 5, Singapore 554910
[c] Sr. Staff Engineer of Process Development Engineering, Kulicke & Soffa, 6 Serangoon North Ave, 554910 Singapore
[d] Advanced Process Engineer of Process Development Engineering, Kulicke & Soffa, 6 Serangoon North Ave 5, Singapore 554910

Abstract

The destructive bond pull test is the most common methodology for quality control of the wire bond process on assembly manufacturing. It was introduced in the 1960's and it is described in U.S military specifications[1)]. However, compared to 40 years ago, the wire bonding process has made rapid progress. Today, some of the hottest processes that System-In-Package (SIP), Package-On-Package (PoP) and 3D package require the Ultra Low Loop (ULL, <75 μm height) process on thinner die (<100μm) [2)]. Therefore, the legacy pull test is not sufficient to apply for the ULL and thinner die applications to meet criteria. Because wire breaking values and the failure mode are significantly different by the loop height, die thickness and the position of the wire pull hook[3)].

This paper is presented in order to introduce a universal wire pull test specification and methodology for the ULL on thinner die application.

Introduction

Today, advanced wire bonding and assembly technology is driving for stacked die, PoP and SIP processes. CAAGR in 2011 will increase by 13% compared to 2006[4)]. The chart below illustrates stacked die and PoP packages.

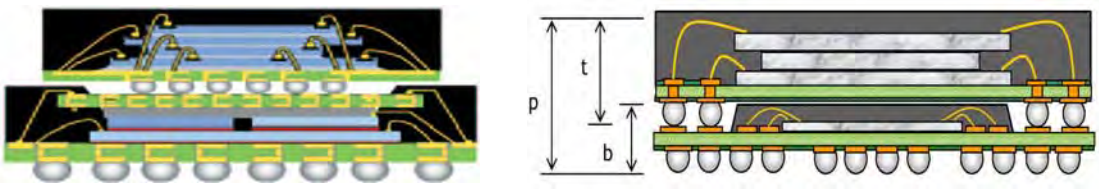


Figure 1: Multi stack and PoP package. *Source: Prismark*[5)]

Ex. 1016, Page 831

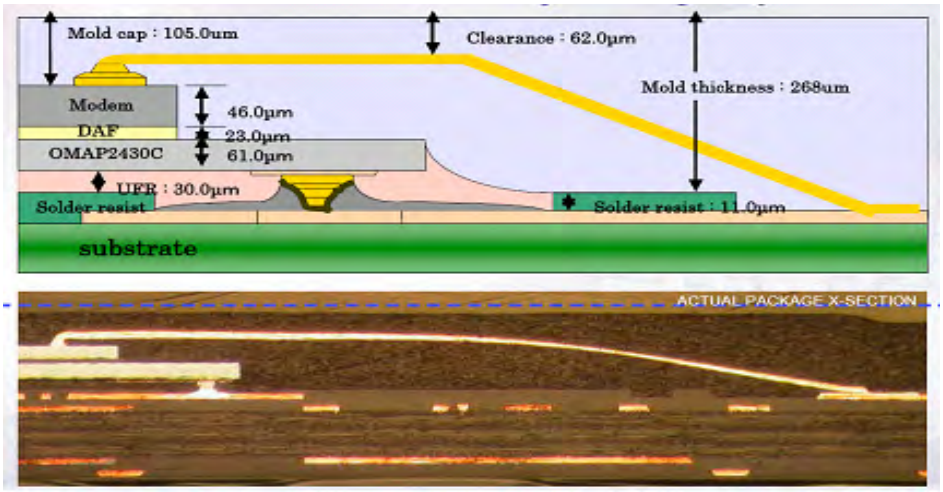


Figure 2: Stack Die Package Thickness and its Looping. *Source: Prismark*[6)]

Package thickness is to 1.0mm for stack die applications. Figure 2 is a 3D package dimension and its cross-section. Wafer thinning, die attach and wire bonding are the key technology for the advanced package. Lower loop height control is the main requirement in the wire bonding process. In 2006, Kulicke & Soffa introduced the less than 60um loop height ULL technology[7)], and then it presented a lower than 50um loop height ULL on even thinner <4mils die thickness on the IConn^{PS} wire bonder in year 2008[8)]. The 50um loop height advanced process is being introduced into mass production successfully with K&S wire bonder.

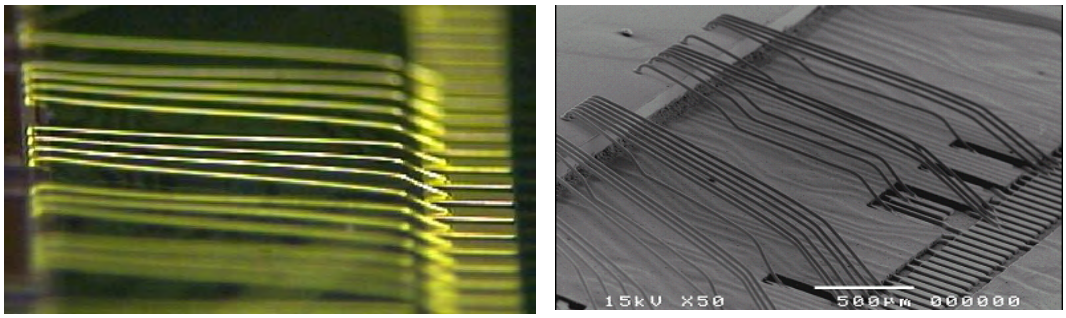


Figure 3: K&S ULL loop shape on various thinner die packages.

This paper will analyze the variables of the wire bond pull force distribution, both theoretically and experimentally, especially for ULL on thinner die processes. It will also find a possible error and problem of the required pull force criteria and Cpk specification in ULL wire bonding quality and reliability.

Equation of the Destructive Wire Bond Full Test

Conventional destructive wire bond full test methodology is well described in the 'Wire Bonding in Microelectronics Materials, process, Reliability, and Yield'[9)]. This is a theoretical model used to demonstrate the geometrical configurations of the wire loop as well as the several equations that define the force distribution. The loop height, the wire length, the position of the pull hook, the height between the pad and 2nd bond surface are all factors influencing the test result.

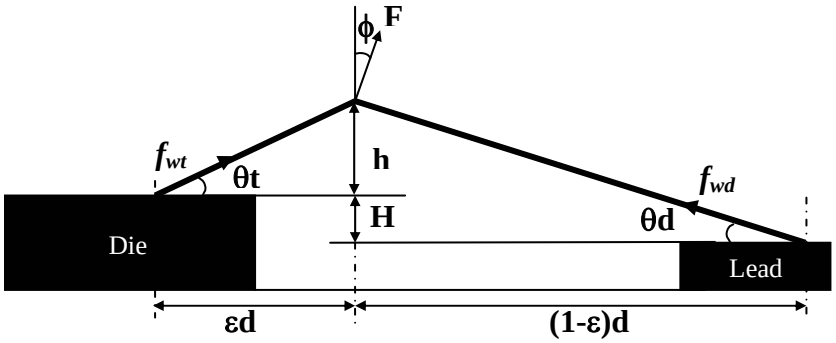


Figure 4: Geometric variables for the test

The force in each wire, (f_{wt} and f_{wd}), at break, with a specified pull force, F, at the hook is:

$$f_{wt} = F\left[\frac{(h^2+\varepsilon^2 d^2)^{1/2}\left((1-\varepsilon)\cos\phi + \frac{(h+H)}{d}\sin\phi\right)}{h+\varepsilon H}\right] \quad [1]$$

$$f_{wd} = F\left[\frac{\left(1+\frac{(1-\varepsilon)^2 d^2}{(H+h)^2}\right)^{1/2}(h+H)\left(\varepsilon\cos\phi - \frac{h}{d}\sin\phi\right)}{h+\varepsilon H}\right] \quad [2]$$

Equivalent equations using angles θ_t, θ_d, and F are:

$$f_{wt} = \frac{F}{\sin\theta_t + \cos\theta_t \tan\theta_d} \quad [3]$$

$$f_{wd} = \frac{F}{\sin\theta_d + \cos\theta_d \tan\theta_t} \quad [4]$$

Note that f_{wt} and f_{wd} are the force or tension in the wire, and in order to calculate the wire pull force, F must be solved in the equation. A wire will break when either f_{wt} or f_{wd} first reaches its breaking strength to each side of the wire. Typically, this is about 90% of the manufacturer-specified breaking load for Au wire (either ball or crescent bond break). A break normally occurs just above the ball in the heat–affected zone. The position of the hook (indicated as εd) and the pull angle, ϕ will significantly affect the distribution of forces at the bonds. One can choose an ε or ϕ value that will give equal forces on each bond, and it will result in more equal test of both bonds. If ϕ = 0 and $f_{wt} = f_{wd}$:

$$f_{wt} = F\left[\frac{(h^2+\varepsilon^2 d^2)^{1/2}(1-\varepsilon)}{h+\varepsilon H}\right] \quad [5]$$

$$f_{wd} = F\left[\varepsilon\frac{\left((h+H)^2+(1-\varepsilon)^2 d^2\right)^{1/2}}{h+\varepsilon H}\right] \quad [6]$$

The hook position to have an equal force of 1st and 2nd bond, $f_{wt} = f_{wd}$ is:

$$\varepsilon = \frac{1}{2+\frac{H}{h}} \quad [7]$$

A plot of the calculated pull force (F) at wire rupture for higher loop height wire bonded on thicker die thickness is given for a typical two-level semiconductor device-bond configuration in Figure 5, pulled straight up $\phi = 0$ at the center of the loop.

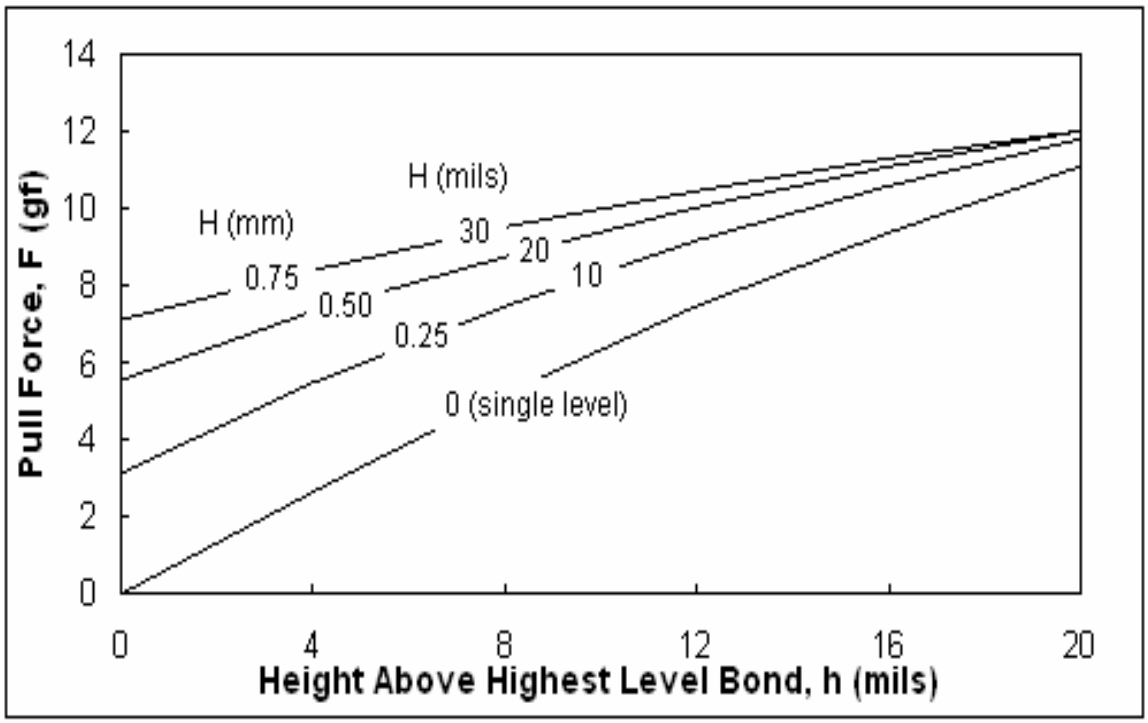


Figure 5: Pull force manner of higher loop height on thicker die thickness. Bond pull force for legacy packages calculated with Eqs. [5] and [6] for various loop heights and package pad heights, pulled in the center of the loop. Where ϕ=0, d=60mils, and fwd = 10gf.

Theoretical Model of the Wire Pull Force for ULL on Thinner Die

Based on the above equations, the wire pull force is predicated to dramatically decrease on ULL less than 4mils loop height on thinner than 4mils die thickness. This is the most common device configuration for the stacked die, PoP and many of the 3D packages. It represents the calculated pull force as a function of hook position for various loop height (h), die thickness (H), and wire length (d) with 0.8mil (20.3μm) diameter Au wire (breaking load: 7.9gm) bonding. The effects of each function are described below.

Effect of loop height and die thickness

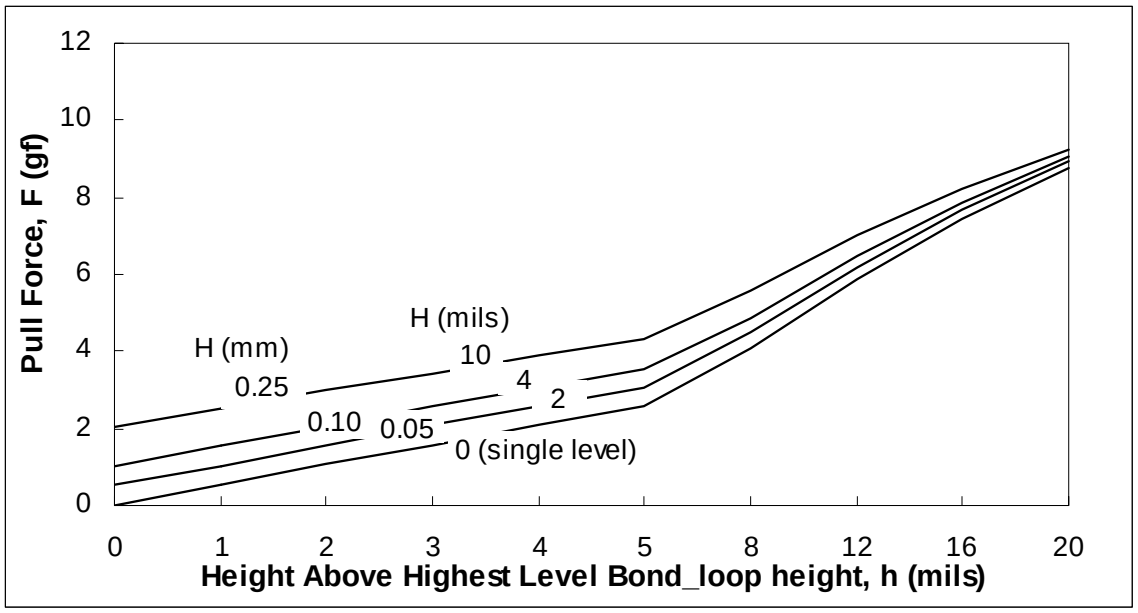


Figure 6: Pull force manner of ULL loop height on thinner die thickness. As seen in Equ 5, the bond pull force is significantly decreased on lower loop height, and thinner die thickness. Where hook position, ε = 0.5 (center of wire length, 30mils from 1st bond), and wire length, d = 60mils.

Effect of loop height with various hook positions

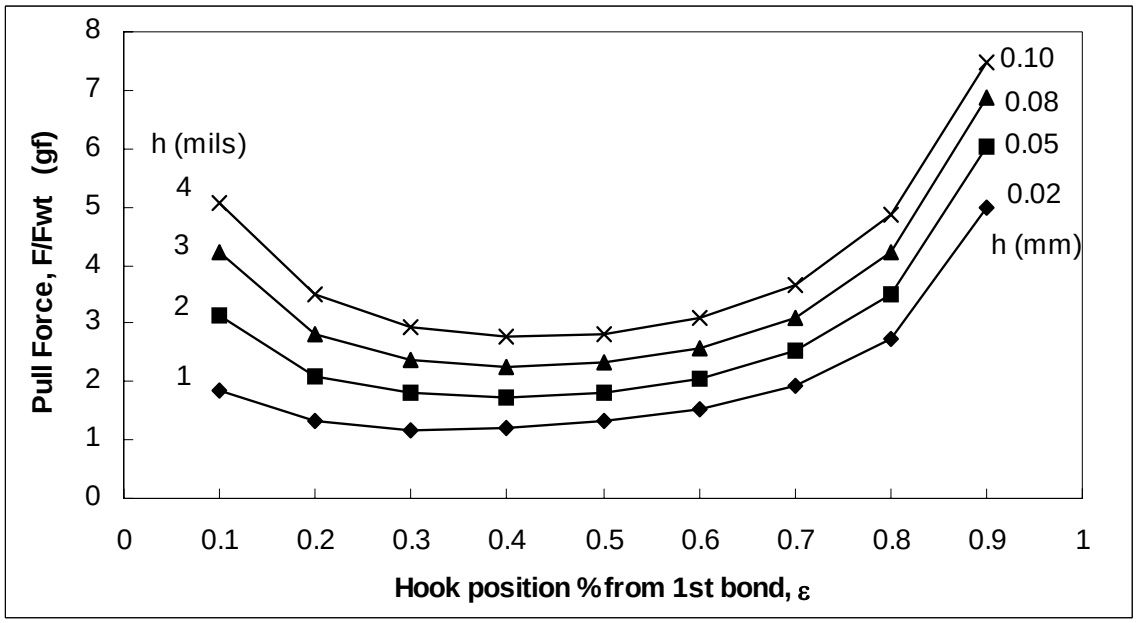


Figure 7: Calculated from Equ. [3]. Bond pull force is significantly decreased on lower loop height. Where die thickness, H = 3mils (0.08mm), and wire length, d = 60mils. First bond is located at ε = 0 and second bond is at ε = 1.

Effect of die thickness with various hook position

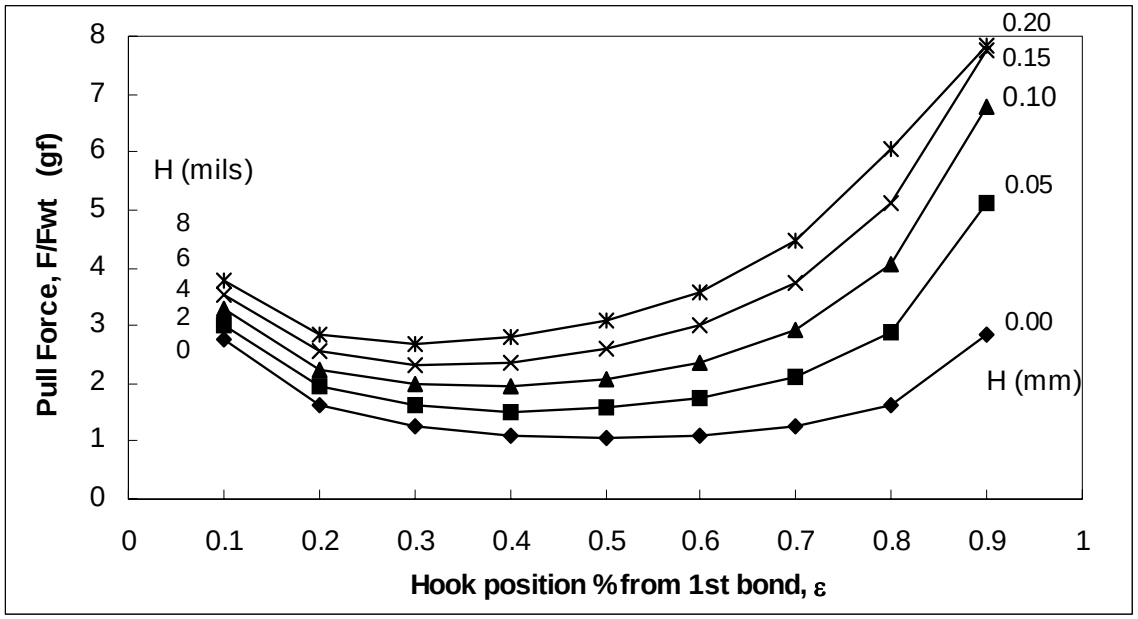


Figure 8: Calculated from Equ. [3]. Bond pull force is significantly decreased on thinner die thickness. Where loop height, h = 3mils (0.08mm), wire length, d = 60mils. First bond is located at ε = 0 and second bond is at ε = 1.

Effect of wire length with various hook position

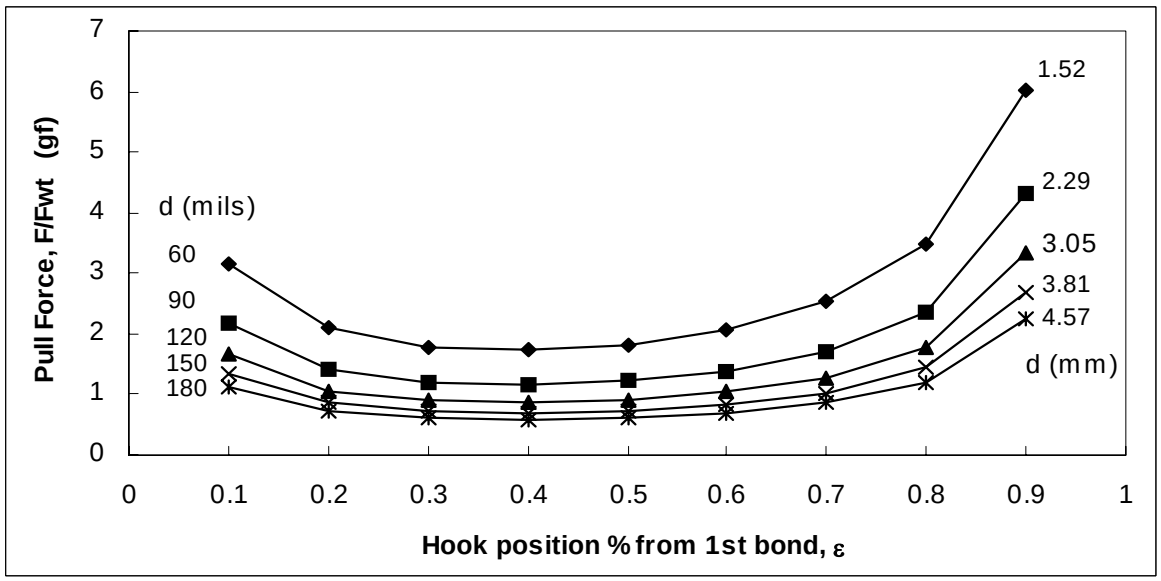


Figure 9: Calculated from Equ. [3]. Bond pull force is significantly decreased on longer wire length. Where loop height, h = 3mils (0.08mm), die thickness, H = 3mils. First bond is located at ε = 0 and second bond is at ε = 1.

Experimental data of the Wire Pull Force for ULL on Thinner Die

As illustrated above on the calculated wire pull force distribution for ULL on thinner die, an actual wire pull force of the new package type like PoP, SIP, and 3D packages would be much lower than most common wire bonding devices. Experimental data will be used to understand difference in the wire pull force as well as process capability (Cpk) for ULL wire bonding on thinner die. There is progress for 3mils and 4mils thinner die thickness and 2mils loop height on K&S IConn^{PS} wire bonder with 0.8mil (20.3μm) Heraeus Formax™ Au wire. On the experiment, it used 2.4gf lower spec limit (LSL) [10] to calculate Cpk, and its target is greater than 1.67 Cpk. The elongation of the wire, the hook diameter, the hook angle impact[11], and pull tester HW setting are not considered in analysis of the experiment data. The additional wire payout to form the last kink loop shape that will affect θ_t, θ_d during pull test is also ignored on the pull force calculation. Therefore the experimental pull force will be a little different compared to the theoretical calculation force.

Before starting the experiment, loop as well as all wire bonding parameters was carefully optimized to remove any abnormal noise. The ball neck and loop shape followed the normal K&S compressed ULL process, and a complete pre-inspection was conducted to ensure there was no ball neck damage and stitch heel crack during wire bonding with SEM. Refer the Figure 10.

Wire pull was executed with the Dage pull tester series 4000-WP100 cartridge. A test load of 3g, speed of 635μm/sec, and hook diameter of 2mils were used.

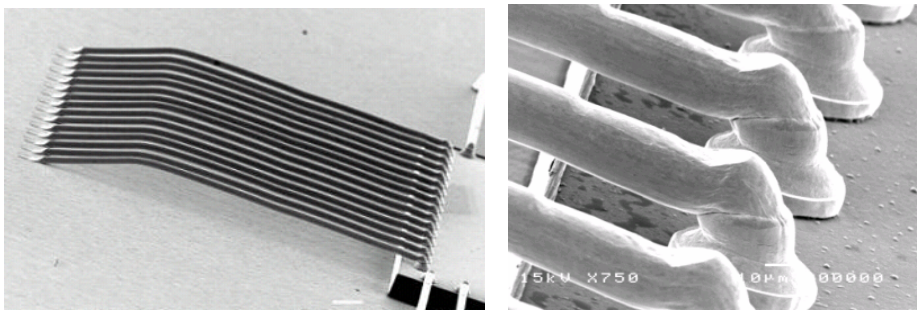


Figure 10: Basic loop shape and compressed ball neck formation for 2mils loop height on 3mils die.

Wire Pull Force of ULL process on thinner die thickness

Figures 11 and 12 are the wire pull force and Cpk for 2mils loop height ULL with various wire lengths and hook positions on 3mils and 4mils die thickness applications. As confirmed by the calculated wire pull force model, the pull strength is increased with higher die thickness (H), the closer the hook positions to 1st bond, and the shorter the wire length. Thus, the force of the ULL will be much lower compared with conventional die thickness (>5mils) and loop height (>4mils) device. Therefore, in order to meet greater than average 3g, the general requirement to achieve confident SPC, the hook position has to move to close to 1st bond.

Ex. 1016, Page 837

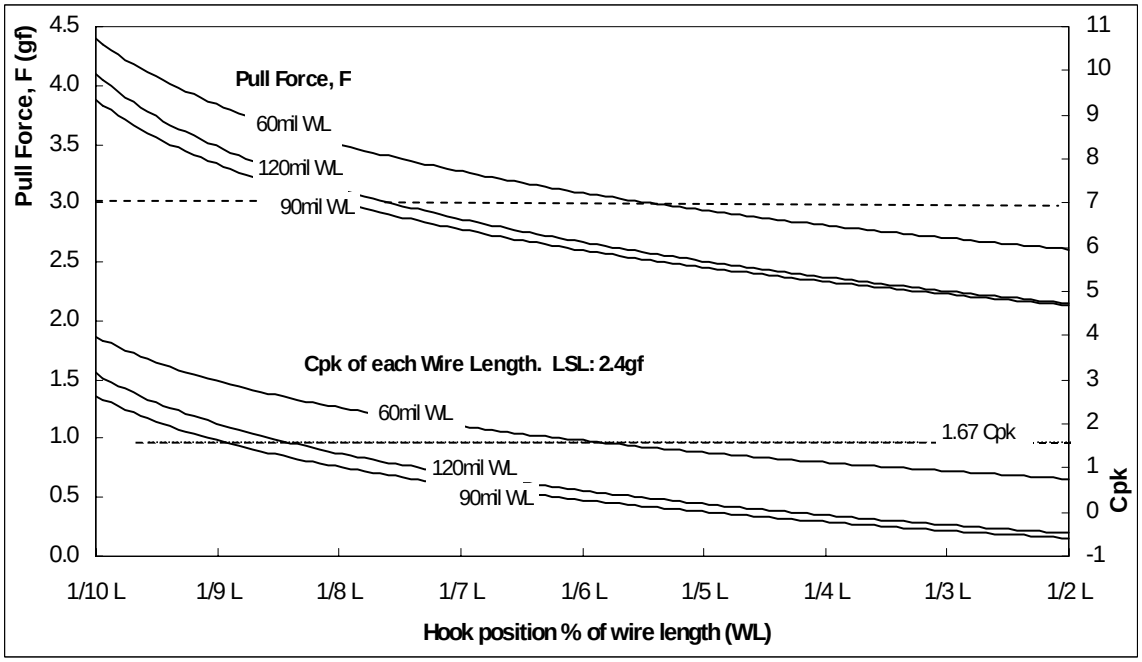


Figure 11: 3mils (0.08mm) die thickness application. Logarithmic trend of the measured pull force and calculated Cpk as a various hook position for different wire length. All of the wire is broken above the ball neck. 0.8mil (20.3µm) Au wire, loop height, h = 2mils (0.05mm). The coefficient of determination, R^2, is >0.8 indicating an acceptable model to describe the variations in, y, pull force.

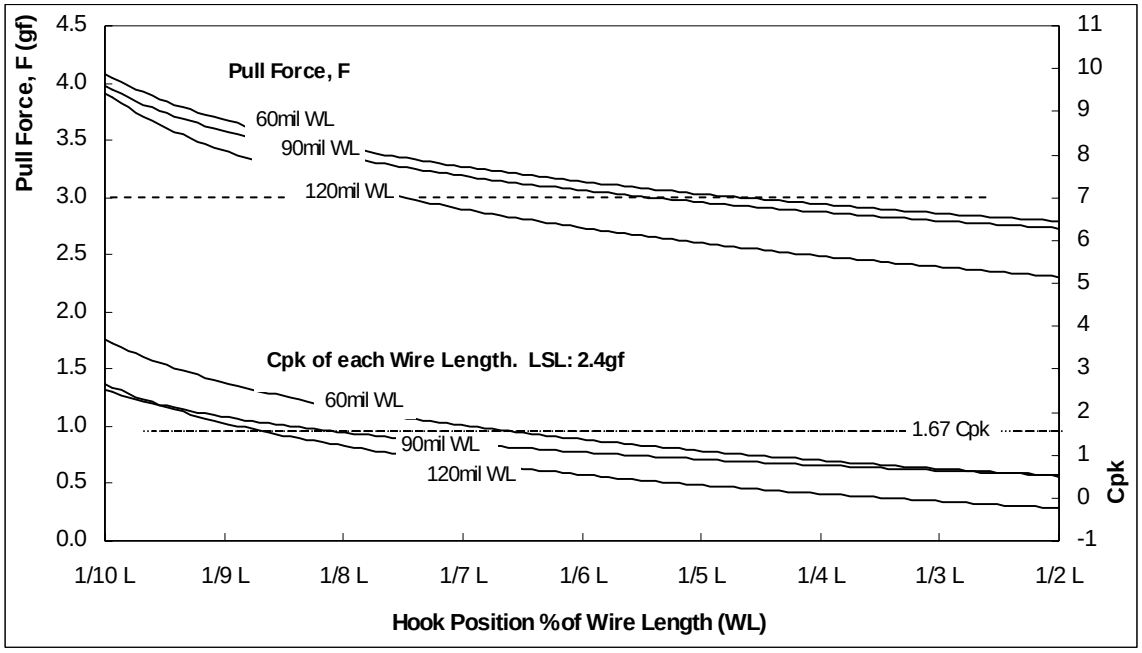


Figure 12: 4mils (0.10mm) die thickness application. Logarithmic trend of the measured pull force and calculated Cpk as a various hook position for different wire length. All of the wire is broken above the ball neck. 0.8mil (20.3µm) Au wire, loop height, h = 2mils (0.05mm). The coefficient of determination, R^2, is >0.8 indicating an acceptable model to describe the variations in, y, pull force.

Cpk is dramatically increased as the hook position is moved closer to the 1st bond location. For the 120mils wire length on the above thinner die, the hook must be located closer than 1/8 L of wire length in order to meet the 1.67Cpk specification. For thinner die and/or lower loop height, and/or longer wire length configurations, the hook needs to move even closer to the 1st bond location. The universal wire pull test method for the SPC and quality specification like Cpk calculations is based on ⅓ L or ¼ L as given in Equ [7]. Based on the experiment data, this method will fail the Cpk spec as shown in Tables 1 and 2.

TABLE 1. Wire pull force data at ⅓ L of wire length hook position ($\varepsilon = 0.33$).

Die Thickness	**3mils**			**4mils**		
Wire Length	Pull statistical	Min - Max	Cpk	Pull statistical	Min - Max	Cpk
60mils	2.56 +/- 0.11	2.41 – 2.76	0.494	2.79 +/- 0.22	2.23 – 3.27	0.597
90mils	2.15 +/- 0.10	1.88 – 2.31	-0.887	2.71 +/- 0.16	2.40 – 3.06	0.670
120mils	2.18 +/- 0.13	1.98 – 2.39	-0.544	2.62 +/- 0.17	2.25 – 3.02	0.436

Table 1: Wire pull force data at ⅓ L of wire length hook position ($\varepsilon = 0.33$) as the common production requirement. Both average force and Cpk failed to meet the specification.

TABLE 2. Wire pull force data at ¼ L of wire length hook position ($\varepsilon = 0.25$).

Die Thickness	**4mils**		
Wire Length	**Pull statistical**	**Min - Max**	**Cpk**
60mils	2.93 +/- 0.26	2.51 – 3.41	0.673
90mils	2.88 +/- 0.21	2.54 – 3.24	0.733
120mils	2.52 +/- 0.17	2.15 – 2.92	0.244

Table 2: Wire pull force data at ¼ L of wire length hook position ($\varepsilon = 0.25$) as the given calculation in Equ. [7]. Both average force and Cpk failed to meet the specification.

These results indicate that the low pull force reading and its Cpk calculation of ULL on thinner die do not come from the ball neck crack during wire bonding. This matter, as a result of the package's geometrical configuration, cannot be avoided in ULL on thinner die case as is defined by the theoretical and experimental analysis. In order to meet the Cpk requirement, the measurement method should be reviewed by either the wire pull specification (LSL) or the hook position.

Conclusions

The rapidly growing demand for multi stacked die, PoP and 3D packages has prompted the development of new equipment, material and processes in order to meet these needs. Ultra Low Loop (ULL) on thinner die is the wire bonding technology needed to realize a thin and light package size for advanced packaging. An accurate concept of this wire bonding process along with the appropriate quality control needs to be considered. As demonstrated in the above discussion, the current industry pull test methodology is not able to meet the Cpk and process quality requirements. The problem is inherent in the thinner die configuration instead of the ULL wire bonding issue. A new wire pull test methodology and ULL process specification needs to be developed in order to continue to keep up with the market demands.

Ex. 1016, Page 839

Acknowledgments

The authors would like to acknowledge Frank Keller (Kulicke & Soffa, FW) who prepared the experiment device, Jason Kam (Heraeus SG) who supported the Au wire for the testing, and the Kulicke & Soffa SG process development team, led by Oranna Yauw, who contributed their time and effort to support this research. A special acknowledgement also goes to Steve Babinetz (Kulicke & Soffa, FW) and Romeo Olida (Kulicke & Soffa, FW) for sharing their valuable knowledge on ULL developments.

References

1. MIL-STD-883G, METHOD 2011.7, *Bond Strength (Destructive Bond Pull Test)*, March 1989
2. E. Jan Vardaman, *Advanced Packaging Interconnect Trends and Technology Developments*, K&S Interconnect Technology Symposium, 2007
3. Lee Levine, *Understanding the Wire Bond Pull Test*, IMAPS 2005
4. Prismark presentation for Kulicke & Soffa, PRISMARK Partners LLC_130 Main Street, Cold Spring Harbour, NY 11724, USA, 2007
5. Prismark presentation for Kulicke & Soffa, PRISMARK Partners LLC_130 Main Street, Cold Spring Harbour, NY 11724, USA, 2007
6. Prismark presentation for Kulicke & Soffa, PRISMARK Partners LLC_130 Main Street, Cold Spring Harbour, NY 11724, USA, 2008
7. Bob Chylak, Stephen Babinetz, Lee Levine, O Dal Kwon, *New Wire Bonding Loops for Stacked Die Packages*, IMAPS 2006
8. O Dal Kwon, *K&S Wire Bonding Looping Developments*, Semicon CEO Talk, In conjunction with Semicon Singapore 2008.
9. George G. Harman, *Wire Bonding in Microelectronics: Materials, Processes, Reliability, and Yield, 2nd ed*, P68~70, New York (1989)
10. MIL-STD-883G, METHOD 2011.7, *Bond Strength (Destructive Bond Pull Test)*, March 1989
11. Shankara K. Prasad, *Advanced Wirebond Interconnection Technology*, p.494~497, Kluwer Academic Publishers group, New York 2004

ECS Transactions, 18 (1) 813-820 (2009)
10.1149/1.3096540 © The Electrochemical Society

Grooved Ball Pad Structure (GBPS) to Improve FBGA Solder Joint Reliability (SJR) Performance

ZHOU Wei, Tom JIANG
Assembly R&D, Micron Technology, Inc.

In this paper, a novel concept using the circular grooved ball pad structure (GBPS) for FBGA is proposed. The goal of the new design is to significantly improve FBGA solder joint reliability (SJR), with minimal impact in cost and packaging assembly processes. Finite element analysis (FEA) tools are used to understand the package solder joint failure mechanism. Simulation shows the GBPS can improve FBGA package SJR significantly. The leadframe based ball grid array package (LFBGA) is chosen as the vehicle to demonstrate the concept of grooved ball pad structure. The reliability results show that GBPS can improve LFBGA package SJR cycles by 2X. Grooved ball pad technology is further attempted on wafer level chip scale packages (WLCSP).

Introduction

Fine-pitch ball grid array (FBGA) package is becoming a main stream package solution for semiconductor devices due to its small form factor and better electrical performance as compared to conventional leaded TSOP packages. However, due to the significant mismatch in the coefficient of thermal expansion between the package and the PCB module board, there has always been a challenge at the module-level solder joint reliability test. The solder joint may encounter cracks or even opening failures when the package undergoes thermal excursion processes. Underfill process is usually an alternative solution in this circumstance to enhance module-level solder joint reliability performance. However, the underfill process may incur a higher packaging cost and also it makes the packaging rework process very difficult.

Fig. 1 shows the structure scheme of a Leadframe-substrate ball grid array (LFBGA) package. During the initial development cycle, the package was found with poor module-level solder joint reliability performance. The package failed at module-level T/C_C 25 cycles and T/C_A 500 cycles respectively. Fig. 2 shows the cross section of a failed solder joint at T/C_A 500 cycles. A fully propagated crack occurs at the interface between solder bulk and ball pad surface and that leads to the fatal electrical contact failures. In order to overcome such a bottleneck, great effort has initially been put by team to improve package solder joint reliability performance. Design optimizations were carried out to select the best combination of package component's dimensions and their material properties to minimize the stress on solder ball. Process optimizations were implemented too to create a best possible clean ball pad surface for a good ball-to-pad wetting. Unfortunately, all these efforts have been found with limit impact on improving package SJR performance. Researchers noted that the LFBGA pkg has inferior solder joint reliability by nature [1, 2, 3, 4]. Hence a generic and more efficient method has to be explored in order to significantly improve LFBGA package solder joint reliability performance.

Ex. 1016, Page 841

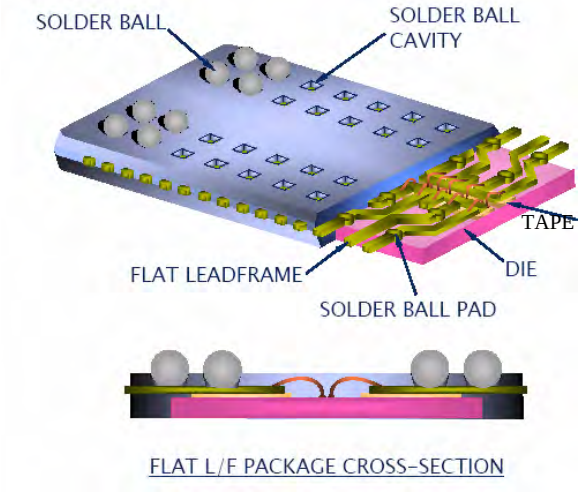


Fig. 1 Structure scheme of LFBGA package

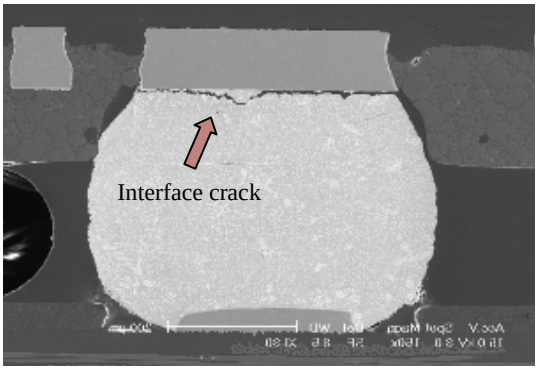
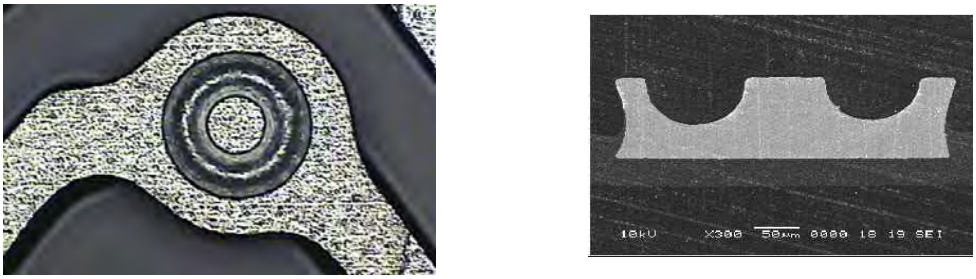


Fig. 2 Cross section of failure units at T/C_A 500 cycles

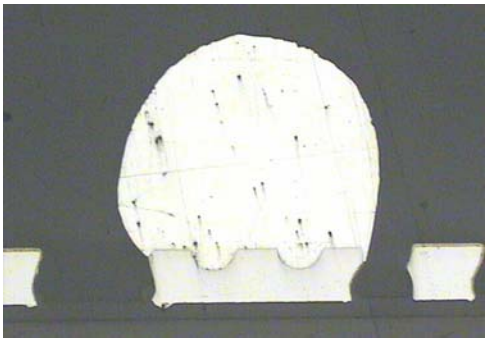
Grooved Ball Pad Structure

A novel concept using circular grooved ball pad structure is proposed here as an answer to significantly improve FBGA solder joint reliability performance. Fig. 3 shows the leadframe samples with grooved ball pad structure. The groove can be designed in any pattern but currently only circular groove is evaluated to simplify the manufacturing process. In future, other types of groove patterns should be explored to maximize the groove benefits.

Ex. 1016, Page 842



(a) (b)

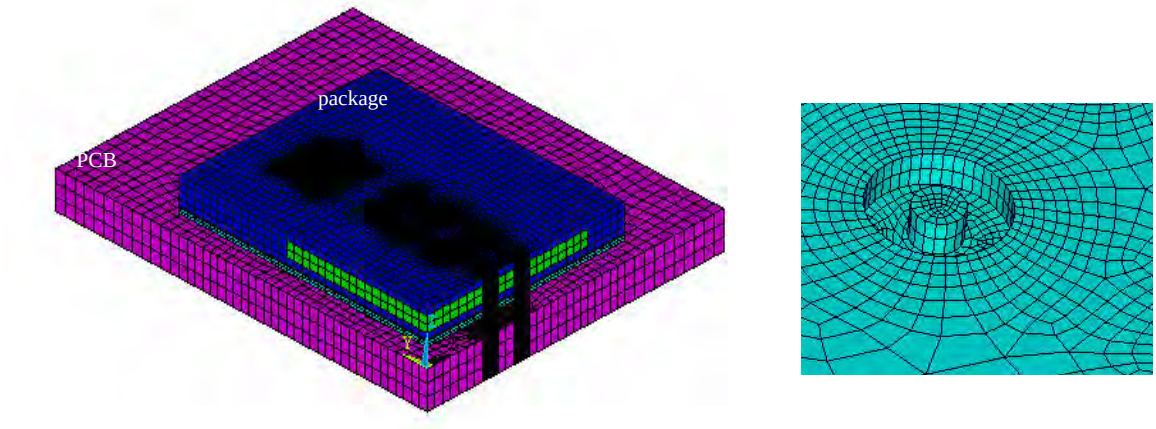


(c)

Fig. 3. Leadframe samples with grooved ball pad structure a) plan view; b) cross section; c) with ball attached.

Finite Element Simulations

In order to minimize the development risks and shorten the development cycle, finite element simulations are carried out in advance to investigate the impact of the grooved ball pad structure on package solder joint reliability performance. Fig. 4 shows the one-quarter model of a LFBGA pkg mounted on PCB.



(a) (b)

Fig. 4 FEM model of a LFBGA package (a) and GBPS (b)

Simulation found that the grooved ball pad structure is able to reduce the stress on solder joint substantially. The max Von-Mises stress at the corner ball is reduced from

283.5MPa to 245.3MPa. Stress simulation only considers one-cycle of thermal loading. The groove benefit will be accumulated to be more significant when package is under large number of fatigue cycling environment, eg, T/C.

To further understand the stress diversion mechanism brought along by the groove, a detail study on the interface stress between solder bulk material and pad surface is carried out (Fig. 5). Figs. 6 and 7 are the plots of stress distribution across the interface centre line. Without the groove, the interface stress is very high at the edge and about half at the central region (Fig. 6). However, with the groove on ball pad, the interface stress is 5 times lower at the groove region and high near the central anchor (Fig. 7). In other words, groove has created a stress buffer ring near the edge of ball-pad interface.

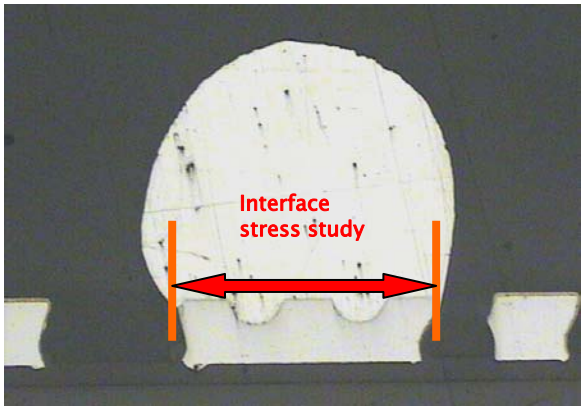


Fig. 5. Interface between solder and ball pad is isolated for interface stress study.

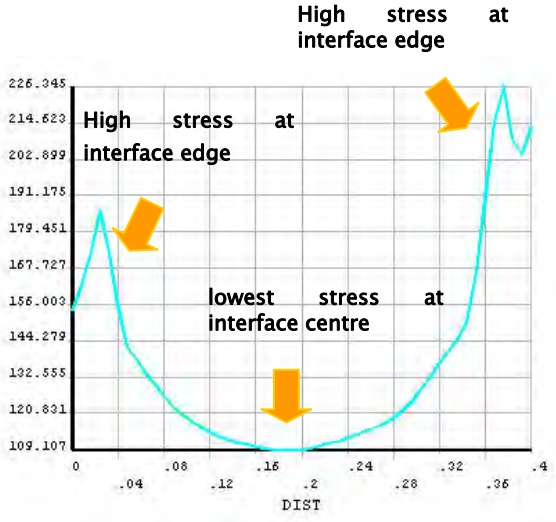


Fig. 6. Interface stress distribution along centre line of a ball pad without groove design

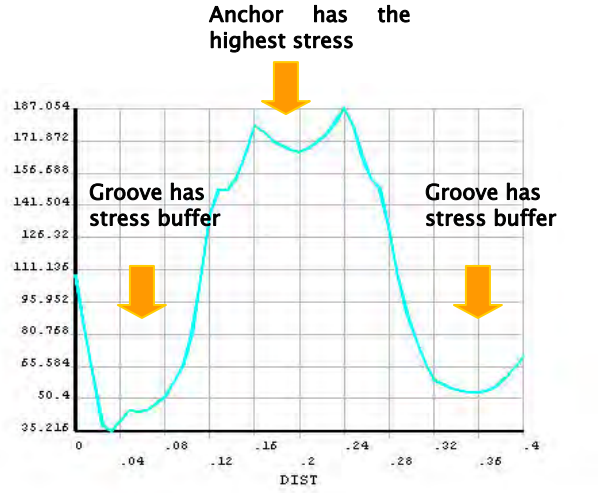
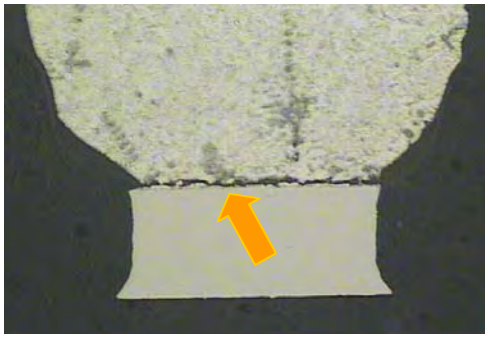
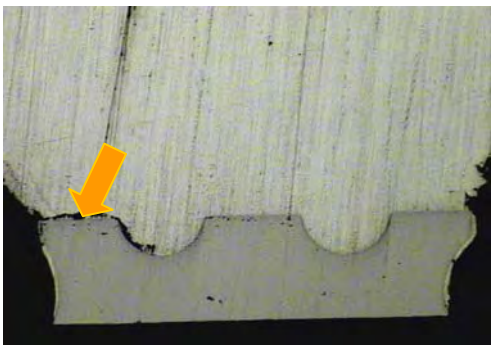


Fig. 7. Interface stress distribution along centre line of a grooved ball pad

On the other hand, according to the generics of fracture mechanics, interface crack, once started, will always be easier to propagate along a flat, straight interface [5, 6]. With the introduction of groove on ball pad, crack will need more energy to propagate along a curved solder/pad interface. This will delay the occurrence of crack full propagation and thus extend the solder fatigue life. This mechanism is later confirmed by FA findings on the actual samples. Fig. 8 shows the cross sections of interface crack of two kinds of pad-solder interfaces: a) pad without groove at T/C_A 500 cycles and b) pad with groove at 1500 cycles. It clearly demonstrates that interface crack has to take longer time to propagate along the curved groove surface and is buffered by the groove low stress.



(a) Without groove, interface crack propagated fully along the straight interface at T/C_A 500 cycles



(b) With groove, Interface crack goes along the curved surface (T/C_A 1500 cycles)

Fig. 8 Cross sections of pad-solder interface with (b) and without groove design (a)

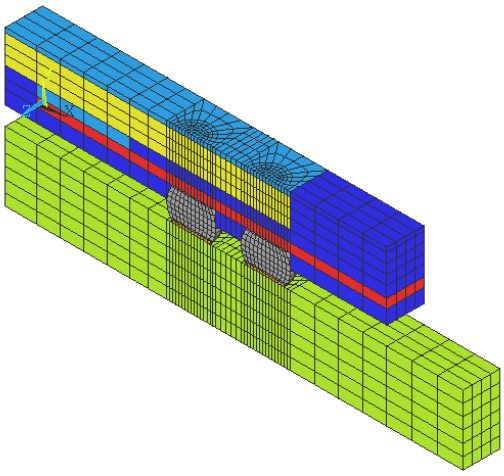


Fig. 9. 3D strip model for solder joint fatigue life prediction

To predict solder fatigue life in thermal cycling environment, a 3D strip modeling is applied to optimize the calculation efficiency and groove 3D effect. Fig. 9 shows the strip model used for SJR prediction. Simulation results show that the grooved ball pad structure is able to improve pkg module level solder joint reliability significantly, from 774 cycles to 1473 cycles.

Sample Reliability Evaluations

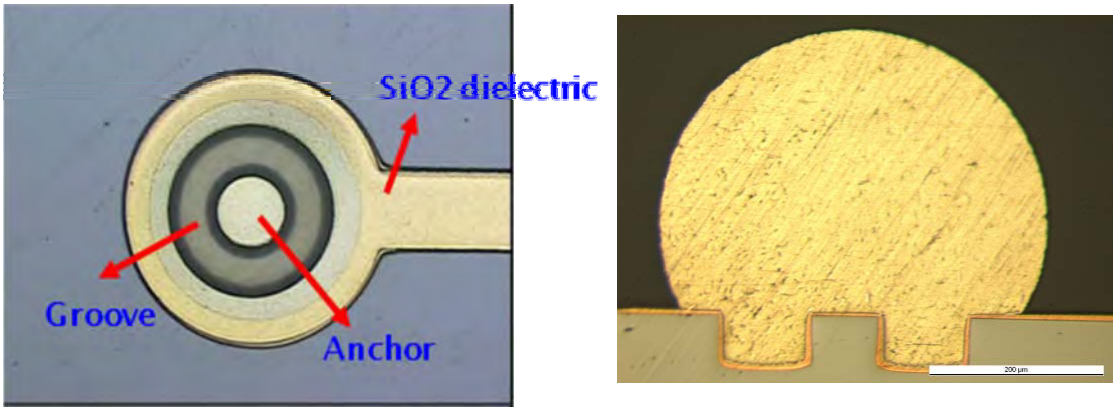
Module level thermal cycling evaluations are carried out. Table I lists the results of failures for T/C_A. Data clearly demonstrates that the grooved ball pad has improved the LFBGA solder joint reliability performance significantly. The improvement on pkg solder joint reliability by groove technology alone is more than 2 times.

Table I. T/C_A module level failures

Assembly	500	750	1k	1.25k	1.5k
LFBGA without grooved ball pad	15/25	10/10			
LFBGA with grooved ball pad	0/20	0/20	0/20	5/20	15/15

Applications to WLCSP

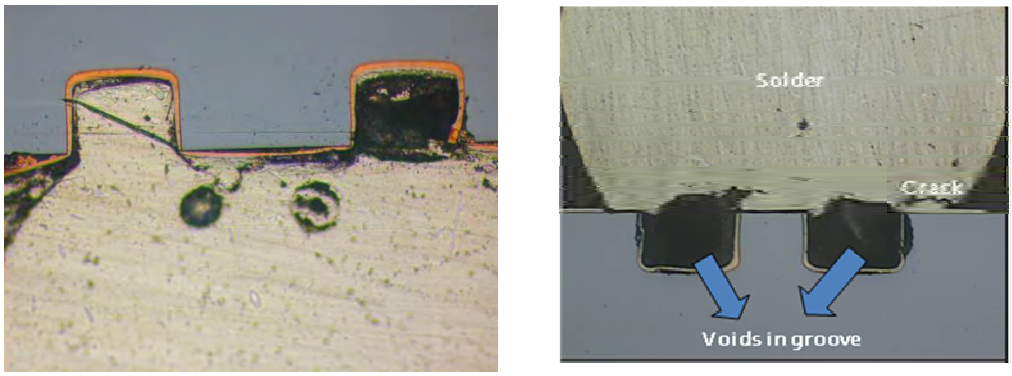
With the success in developing groove technology on LFBGA package, the grooved ball pad structure (GBPS) is further attempted to wafer level chip scale package (WLCSP) where no substrate is present. The SJR performance for WLCSP is even more critical as silicon has much larger CTE mismatch with PCB as compared with FBGA packages. Fig. 10 shows the creation of groove on the silicon surface before and after ball attachment.



(a) (b)

Fig. 10. Groove ball pad structure for WLCSP before (a) and after ball attachment (b)

However, there are a few assembly challenges related with GBPS for WLCSP. The main issue is that the flux is easily entrapped in the groove and becomes voids during ball attachment process. Fig. 11 shows the cross section of voids entrapped in the groove. Measures are being explored to improve the groove surface wetting ability and look for ways to facilitate the volatile escaping. The other challenge is the delamination along the solder bulk and the groove metal wall surface (Fig. 12). This may be caused by the contamination or oxidation along the groove wall surface. Investigations are still ongoing.



(a) (b)

Fig. 11. Voids in the GBPS for WLCSP. (a) partially and (b) completely

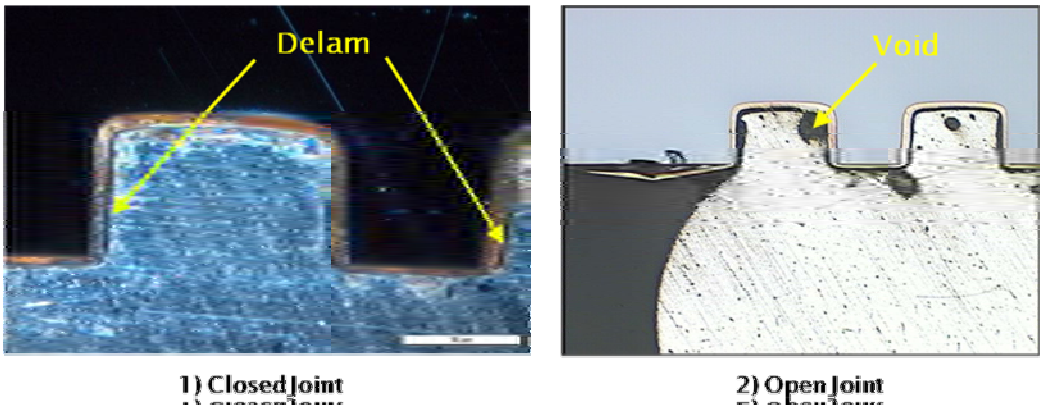


Fig. 12. Delaminations found along the solder bulk and groove wall surface. (1) closed joint and (2) open joint

Conclusions

The grooved ball pad structure is proven by both finite element simulations and sample evaluations that it can improve LFBGA sold joint reliability significantly. Hence. Underfill can be eliminated from the assembly processes. Study shows groove reduces the pad/solder interface stresses by creating a stress buffer ring near the critical interface edge. In addition, the groove curved pad/solder interface extends the length of crack propagation path and thus prolongs the solder joint fatigue life. GBPS technology was attempted on WLCSP too. Initial evaluation encountered entrapped voids in the groove. Counting measures to eliminate the voids in the groove are explored and the void mechanisms are also investigated.

References

[1] Zhou Wei, Lim Beng Kuan. Enhancing solder joint fatigue life for ball grid array packages. *EPTC, 2003.*

[2] J. T. Moon, S.H. Hong, Seung Wook Yoon, Y. H. Choi, Ik Seung Park. Fabrication process of copper lead frame chip scale packages. *ECTC, 1999.*

[3] Bor Zen Hong, Lo Soun Su. On thermal stresses and reliability of a PBGA chip scale package. *ECTC, 1998.*

[4] M. Amagai, M. Nako. Ball grid array packages with the copper core solder balls. *ECTC, 1998.*

[5] Moran, B. and Shih, C. F. crack tip and associated domain integrals from momentum and energy balance. a general treatment of crack tip contour integrals. *Engineering Fracture Mechanics. Vol. 21, pp.405-421. 1985.*

[6] Sih, G. C. Introductory chapter: A special theory of crack propagation. *Mechanics of Fracture Vol. 1, pp.1-55. 1973.*

ECS Transactions, 18 (1) 821-830 (2009)
10.1149/1.3096541 © The Electrochemical Society

Microdot Epoxy Stamping Process Qualification for Substrate-Based Optocoupler BGA Packaging

Emmanuel Livelo, R'veen Pierangelo V Rojas
Fairchild Semiconductor
emmanuel.livelo@fairchildsemi.com, r'veen.pierangelo.v.rojas@fairchildsemi.com

Optocoupler is a semiconductor device that allows signal to be transferred between circuits, while keeping the circuits electrically isolated from each other. The demand of surface mount Optocoupler packaging has progressed from a conventional Dual-in-line packaging (DIP) to Small outline packaging (SOP) to miniflat packaging (MFP) and to Ball Grid Array Packaging (BGA) for Optocoupler. This innovative design employs the use of premolded lead frame technology offering thinner and smaller package structure. The need to maintain a consistent interconnect dispense material at die attach is very critical to attain a reliable adhesion of the epoxy to the die and to the substrate. Stamping or pin transfer approach was qualified as a die attach dispense system. This paper outlines the overall characterization work in the die attach process to meet the packaging requirement of the substrate based Optocoupler in a BGA package.

1. INTRODUCTION

Microcoupler is Fairchild's version of a optocoupler device that uses a short optical transmission path to transfer a signal between elements of a circuit, typically a transmitter and a receiver, while keeping them electrically isolated in a BGA packaging footprint. Below illustration shows Fairchild's Optocoupler device:

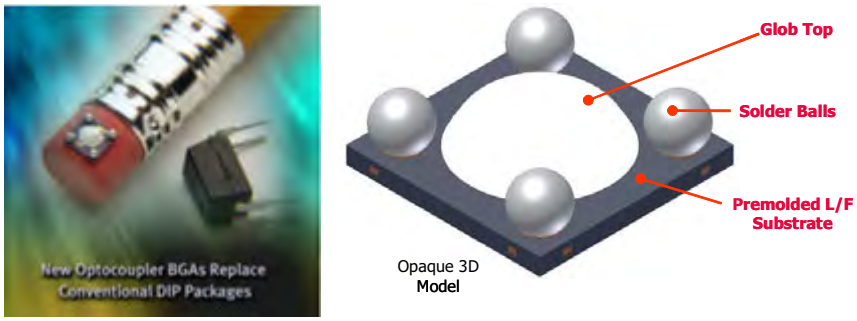


Figure 1.0 Fairchild's Optocoupler in a BGA packaging

The advantages offered by this optocoupler package include smaller size (reduced footprint and thickness), lesser weight and improvement in CTR performance over temperature. It has a wide operating temperature range as compared to conventional optocouplers. Owing to its low profile, small volume and footprint, this package can further save on board space giving hardware designers more flexibility and allowing for overall systems cost savings. Optocoupler BGA packaging poses great challenges in large scale manufacturing. Due to the substrate density and smaller pad size of the frame, major challenge in the manufacturing is to maintain minimal warpage and pad contamination of substrate leadframe. Characterization in the selection of molding compound and tape material was done in parallel with frame design optimization to

Ex. 1016, Page 849

reduce warpage. Substrate and plasma cleaning before die attach process further guarantees removal of pad contamination problem before the die attach process.

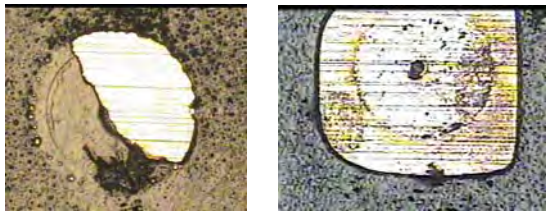


Figure 2.0 Example of pad contamination problem (flash and residues)

In the die process, substrate warpage is critical is indexing, dispensing and bonding processes. Proper design of clamping system will ensure planarity of the substrate for stamping and die attach. Maintaining a consistent volume of epoxy volume for die attachment was achieved by implementing stamping approach among other epoxy dispensing technologies. Epoxy stamping tools can stamp various patterns of epoxy on the pad, depending on the needs of the application. A dot matrix pattern, for instance, ensures that a consistent amount of epoxy is deposited on the die pad every time. Through stamping, consistent volume of epoxy dispensed was achieved. The internal isolation gap in between the LED and photo detector component is critical for the safe operation of the device. The isolation distance must be free from any metallic contamination thus keeping the epoxy within the boundaries of the bonding pad is critical at the Die Attach process.

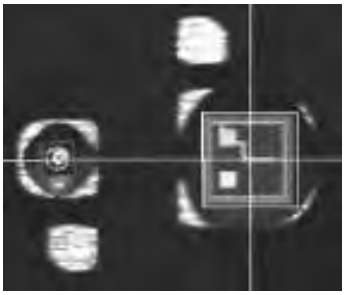


Figure 3.0: Screen shot of the LED and Photo detector (transistor) Component of a Microcoupler at the Die Attach Process

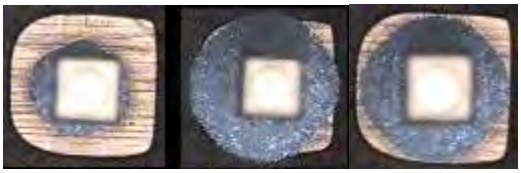


Figure 4.0 Illustration of an insufficient, excessive and ideal epoxy volume for the LED component

1.1 Objective

The scope of the work is on the die attach process. Characterization has been performed to attain a consistent microdot volume of epoxy material for the LED and photo detector component while keeping the isolation area free from contamination and attaining a

Ex. 1016, Page 850

consistent epoxy fillet for the LED. Equipment used in the characterization study is three in one modular equipment with epoxy stamping capability.

1.2 Basic Adhesive Application Process

In general, dispensing can be defined as transfer of fluid from a container (usually a syringe) to a substrate (a leadframe) in a form of pattern with calculated volumes to host a die or component. In a semiconductor industry, various methods have been evolving in order to keep up with the rapid development of manufacturing need for adhesives in a wide array of packaging assemblies. Traditionally, there are three known methods of applying Epoxy into a substrate DAP (1) Syringe Dispensing (2) Screen Printing (3) Stamping (Pin Transfer). Stamping approach or pin transfer approach was the best evaluated option in meeting the requirement for microdot application on a challenging premolded substrate condition.

Dispensing process was earlier tested on time pressure and auger systems. The attempt to used these technologies failed due to either of the following (1) work holder design of the equipment does not allow proper clamping of the leadframe resulting to non planar condition (2) dispensing module does not allow degree of overtravel to maintain consistent Z height during dispensing to compensate frame warpage/planarity (3) dispensing tool had limited area to provide standoff design to maintain consistent Z height during dispensing. (4) Time pressure equipment has no capability for dispensing at micro volume. With these constraints, the stamping approach was evaluated as an alternate solution for microdot dispensing application. Screen printing process is not usually used for creating microdots on a substrate pad leadframe. Design of a work holder and stencil design for this application needs additional amount of engineering investment.

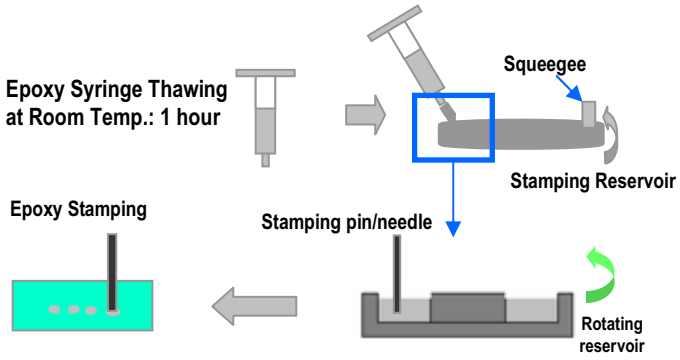


Figure 5.0 Illustration of an Epoxy Stamping process

1.3 Adhesive Stamping Process – Process Definition

Pin transfer or stamping process employs the following components: (1) a rotating reservoir maintaining a defined level of epoxy. The level of epoxy is controlled by the squeegee level which can be mechanically adjusted. (2) Stamping module with Z height over travel capability. For this equipment used in the study, the stamping module is a retrofitted bonding head with this capability. (3) Transfer pin which comes in contact

Ex. 1016, Page 851

with the epoxy during dipping and comes in contact with the bonding pad during epoxy transfer. (4) A tool holder to cater switching of pin transfer tools for LED and transistor.

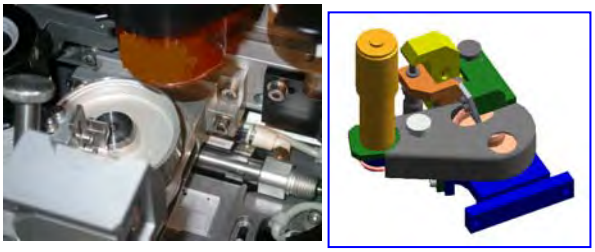


Figure 6.0 Actual and Schematic Drawing of the Epoxy Stamping module

In the epoxy stamping process, the epoxy is held in a rotating reservoir holder. The height or depth of the epoxy level inside the reservoir is held constant through a squeegee. The rotation speed is user programmable. During stamping process, the pin is dipped into the epoxy at a pre defined depth and the epoxy is pin transferred to the bonding pad by the bonding module. Stamping for the other die component is done through a switching carousel which is auto programmable to switch from one stamping tool to another.



Figure 7.0 Illustration of the Epoxy Pin transfer work area

The equipment used in the qualification is an in-line 3 modular equipment with stamping capability in Module 1 and bonding capability in modules 2 and 3. Each module is equipped with same similar system software and frame handling capabilities. Transport of substrate from one module to another is done through a belt transport system. Securing the substrate and maintaining it planarity during stamping and bonding processes is done through a vacuum held clamping system.

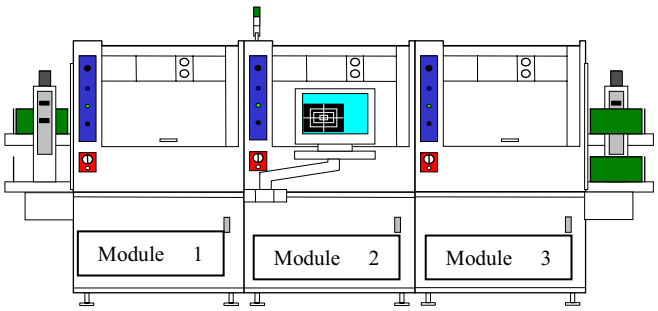


Figure 8.0 Schematic view of the Die Attach Equipment

Ex. 1016, Page 852

The process flow designed in this modular equipment is as stratified in the following matrix. Pre bond inspection after stamping process is employed to guarantee presence of epoxy in the bonding pad and is programmable to do skip bonding for missing epoxy.

Die Attach Sub Process Map	Module 1	Module 2	Module 3
Leadframe Index from Magazine Input			
Bonding Pad Check			
Epoxy Stamping (LED)			
Epoxy Stamping (Photo Detector)			
Pre Bond Epoxy Dot Check			
Bonding Photo Detector			
Die Check for PhotoDetector			
Bonding LED			
Index to Magazine Output			

Table 1.0 Die Attach Sub-Process Map

2.0 EXPERIMENTAL SECTION

2.1 Stamping Tool Design Evaluation

Several tool design concepts were evaluated to attain a consistent epoxy volume and size. The first attempt was to implement a gang type epoxy stamping tool design concept that incorporate both the LED and photo detector tool design in a single holder. This maximizes productivity of the stamping process.

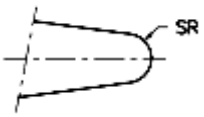
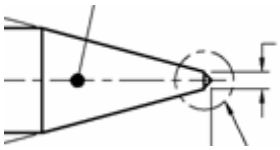
Tool	Tool Design	Tip Configuration / Concept
1	Round Tip	
2	Star Tip	
3	**Chamfered Tip**	

Table 2.0 Stamping Types

The tools were tested at different dipping depth at the reservoir to attain optimum amount of epoxy volume for stamping. The gang stamping tool was tested by doing sample stamping and the epoxy coverage and epoxy fillet responses were assessed by conducting visual inspection at low magnification scope.

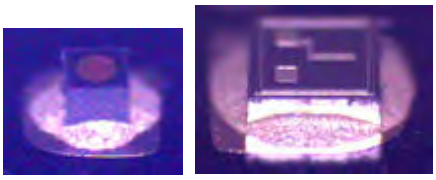


Figure 9.0 Epoxy Fillet Requirement for the LED and Photodetector

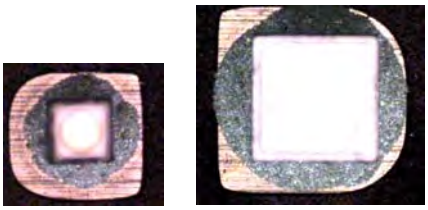


Figure 10.0 Epoxy Coverage for LED and Photo detector die

2.2 Resin Bleed Out Control – LF Staging Characterization

Substrates after pre molding process are subjected to plasma process to remove residual contamination on the bonding pads from the cleaning medium. The plasma system used is a Physical type plasma cleaning system where ions generated in an argon plasma bombard the surface with sufficient energy to sputter contamination from the surface. his process leaves a residue free bonding pads for die attach process. Contact angle measurements lowered the response from 80degrees (uncleaned pad) to 10degrees after implementing plasma process.

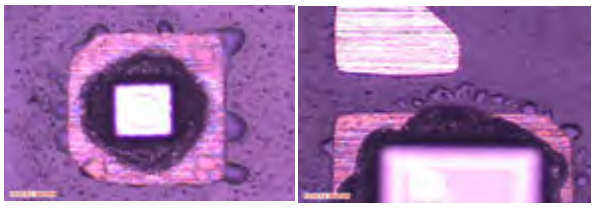


Figure 11.0 Resin Bleed out after stamping and Die attach

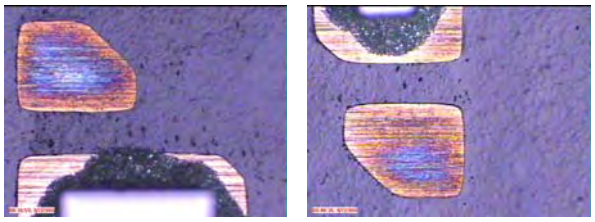


Figure 12.0 Residue from Resin Bleed after Curing

However, during stamping process, visual inspection of the bonding pads showed resin bleed out phenomenon or RBO, where the lower polymeric component of the epoxy resin migrates aong the surface of the substrate by capillary action. This phenomenon resulted to contamination of the adjacent bonding pad for wirebond area. After die attach curing, this left a residual contamination that results to bondability issues at wirebond process.

Ex. 1016, Page 854

The bleed out resin dries after curing leaving behind a stained like residue on the weld pad area detectable only a high magnification scope visual inspection. The defect is concentrated largely on both ends of the substrates which is the exit flow of the cured solvents during curing. Spacing of substrate improves out gassing circulation and reduces defect rate but have not eliminated the defects. The stained residue after curing process cannot be removed by the subjecting the cured substrates back to plasma process. Excessive epoxy dispensed correlates with increasing defect rate of resin bleed out. So an optimum amount of epoxy volume is characterized. Characterization of the plasma parameters and staging time before stamping process was done to arrive at a surface pad condition that will not result of resin bleed out problem.

3.0 RESULTS AND DISCUSSION

3.1 Stamping Tool Design Evaluation

The epoxy coverage of different stamping tool design concepts at different stamping levels is summarized in below table.

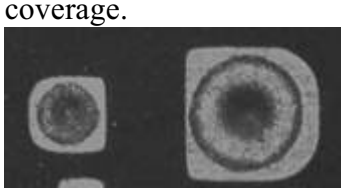
Tool	Tool Design Concept	Sample epoxy Coverage from Gang Tool Design
1	**Round Tip**	Epoxy coverage signature is flat and too wide. This resulted to poor fillet formation on the LED and Transistor during bonding. 
2	**Star Tip**	Acceptable epoxy signature starting at 15um Z depth and excessive epoxy coverage at Z depth at 30um. But still minimal volume for transistor component at an excessive epoxy level for LED at higher Z depth. 
3	**Chamfered Tip**	Similar response with round tip. Coverage is difficult to control to meet both the LED and Transistor requirement due to flat coverage. 

Table 3.0 Stamping Tool Epoxy Coverage Results

In this experiment, it was difficult to achieve optimum volume of epoxy using the gang stamping tool design, both for the LED die and Transistor die, the latter has a bigger die size area than the former. The need to do separate stamping tool for LED and Transistor die is necessary to have individual control of epoxy volume for both components. After re characterization, the rounded tip for LED and star tip design for transistor component has shown to be the most suitable tools design.

3.2 Resin Bleed Out Control – LF Staging Characterization

Since the resin bleed out problem occurred right after stamping process, an evaluation was conducted using different staging time condition after plasma. Visual inspection analysis were done after stamping, die attach and curing process. Contact angle analyses were performed after plasma. Below table summarizes the results. Zero resin bleed out problem were detected after 3 hours of staging after plasma process.

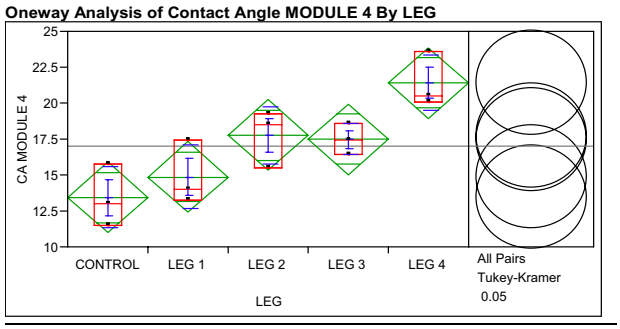
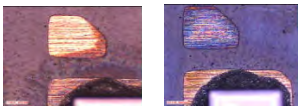
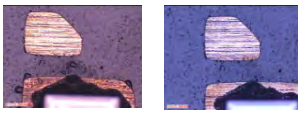


Fig. 13.0 Contact Angle Results

Contact angle analysis on the edge part of the substrate showed increasing contact angle as staging time in increased and doubles the value after 4 hours. Controls were made to implement curing of die attach substrates within 8 hours after plasma cleaning. This is to control the dwell time before cure that will also affect resin bleed out.

Leg	Description	Defect Rate at Sample Size of 200 units
1	Control – Zero Hour staging after Plasma	**17%** 
2	1 hour staging after Plasma	**8.50%** 

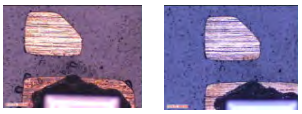
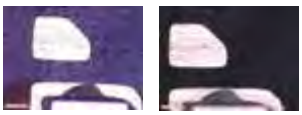
3	2 hour staging after plasma	**1%** 
4	**3 hour staging after plasma**	**Zero ppm** 
5	**4 hour staging after plasma**	**Zero ppm** 

Table 4.0 Substrate Staging Results

4.0 CONCLUSION

An adhesive stamping process was qualified using an in-line modular machine for microdot application on substrate based package for Microcoupler, Characterization of different stamping tool design yielded a rounded tip and star tip design for LED and photo detector die. While implementing plasma cleaning after premolding process yielded clean bonding pads for die attach process, resin bleed out phenomenon was detected due to high surface energy of the bonding pads after plasma. Implementing plasma is necessary to attain clean bonding pads after substrate cleaning, a work in characterizing dwell time after plasma yielded to a staging time of 4 hours before stamping process.

5.0 RECOMMENDATIONS

From the results of the experimentation, Stamping technology can be implemented to achieve a consistent epoxy dot volume for microdot applications, and meeting fillet height and coverage requirements for the Optocoupler dice. It is recommended to use the rounded tip for LED and star tip design for Transistor component as it resulted as the most suitable tools for consistent epoxy fillet coverage. Further work in characterizing rheological characteristic of the adhesive for stamping application is recommended to understand change of adhesive property along cyclical stress and staging time in the stamping process and determine its appropriate exposure life limit for this process.

6.0 ACKNOWLEDGEMENT

The authors would like to acknowledge all the people who have made this project study possible.

Ex. 1016, Page 857

7.0 REFERENCES

1. Cronin, P. and Crowley, D. *et al*, "Advances in Epoxy Dispensing", *Electronic Packaging & Production*, December 2001.
2. Burga, R. and Tausek, A. *"Control of Adhesive Dispensing Parameters during Transition from Research to Production Environments"* EPTC 2003.
3. Properties of Die Attach Adhesive – Resin Bleed Out Phenomena, Ablestik Korea

ECS Transactions, 18 (1) 831-836 (2009)
10.1149/1.3096542 © The Electrochemical Society

Cu^2 Process Development For Cost Effective SOT Package

Soh Yuen Chun and Lim Tian Hing

Infineon Technologies (Wuxi) Co., Ltd, Wuxi Jiangsu, 214028 China

Abstract

Commodity discrete package SOT faces tremendous cost reduction pressure among integrated chip & packaging manufacturers. Low cost per piece is desired to create better gross margin. Cost analysis found that direct materials (leadframe, bonding wire, compound etc) contributed >50% package cost per piece. Therefore, one of the rock bottom ideas will be exploring on bare copper leadframe with copper wire bonding. Current SOT package material sets are NiFe42, silver plated leadframe bonded with gold wire. This could easily reduce leadframe & bonding wire cost by 50%. This paper will discuss details feasibility study & development of Cu^2 (bare Cu leadframe + Cu wire) backend process on SOT. Material sets selection, technical challenges understanding, equipment setup, feasibility study & design of experiment, failure analysis etc are performed to assess process manufacturability. Meanwhile, thermo-mechanical & moisture related reliability assessment (HTS/TC/PCT etc) also will be performed prior to official qualification & customer build.

Introduction

As the gold (Au) price continues on the rising trend, semiconductor packages manufacturers are keen to replace Au wire to cheaper bonding material in effort to reduce packaging cost. Copper (Cu) appears to be a suitable choice; it has certain advantages against Au like cheaper material cost, higher thermal/electrical conductivity, slower intermetallic growth & higher stiffness etc.

Meanwhile, internal analysis reveals that leadframe is also contributing to high packaging cost. Base & plated material determine part of the leadframe selling price. Popular materials in the market are NiFe42 (A42) base with silver (Ag) spot plating. Cu base leadframe seems to be an option. In order to get further cost down, leadframe with non Ag plated is set to be explored.

The objective of this paper is to develop bare Cu leadframe with Cu wire process on SOT to improve cost performances. Details description of activities such as technical challenges, Cu wire kit setup, preliminary trial run, design of experiment, failure analysis, reliability performances etc will be discussed.

Leadframe, Bonding Wire Cost & Properties

Leadframe change from A42 with Ag plated to bare Cu, while bonding wire from

Ex. 1016, Page 859

22um 4N Au to 22um 5N Cu. It has significant cost down on leadframe ~40% & wire ~75%. Highlighted are the main advantages of Cu against Au mechanically & electrically.

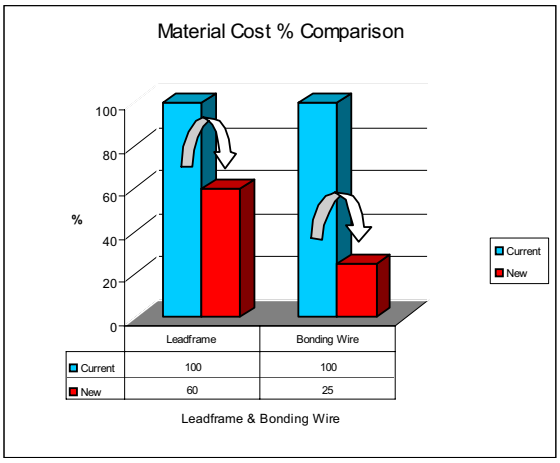


Figure 1. Material cost

Property	Units	Gold	Copper
Melting Point	° C	1063	1083
Density	g/cm^3	19.3	8.92
Lattice Constant (at 20° C)	10^{-10}m	4.079	3.615
Lattice Structure		FCC	FCC
Specific Heat (at 20° C)	J/g K	0.126	0.386
Thermal Conductance	kW/m^{2}K	31.1	39.4
Coefficient of Thermal Expansion	ppm/K	14.2	16.50
Electrical Resistivity (at 20° C)	10^{-8} ohm m	2.2	1.7
Electrical Conductivity (at 20° C)	10^7/ohm m	4.55	5.88
Vickers Hardness	MN/m^2	216	369
Modulus of Elasticity	GPa	78	123
Tensile Strength	N/mm^2	120-220	210-370

Figure 2. Properties comparison

Technical Challenges

Although Cu is a choice to replace current leadframe & bonding materials, it exhibits key challenges in developing & mass production Cu[2] process. As the nature of Cu characteristics, it oxidizes readily under elevated temperature. This creates quality & reliability concern if oxidation is taken place. For example oxidized leadframe may lead to poor adhesion of die attach, lifted or weaken stitch bond, poor adhesion of leadframe to mold compound etc. If oxidation happens at free air ball (FAB) during wire bond, non stick on pad, lifted ball or bond pad crater could occur. Forming gas protection has to be introduced at die attach & wire bond process to prevent oxidation.

Cu wire is harder than Au wire; past literature revealed that Cu FAB is ~30% harder & has ~25% increased of maximum stress on bond pad compare to Au wire bonding. Critical bond parameters need to be optimized properly to prevent pad underlying silicon damage (pad crater), maximizing process yield & reliability performances.

Process Feasibility Study

Equipment Cu kit is needed to handle oxidation free Cu wire bonding. Photo below showed the Cu kit setup for ACBxx wire bonder, it consists of forming gas (95%N_2; 5%H_2) nozzles surrounding EFO & shutter plate at bonding area. Cu FAB oxidation is controlled by nozzles while shutter plate will improve oxidation on leadframe & bonded ball.



Photo 1. Cu wire kit setup.

Initial trial run of Cu bonding was not promising. Ball oxidation, pad cratering & inconsistent of intermetallics was observed. Changes were made to overcome those critical issues. Fine tuning EFO sparking position, slow ramp USG profile, optimum flow rate & high force setting were performed. Photo 2 showed the performances comparison.

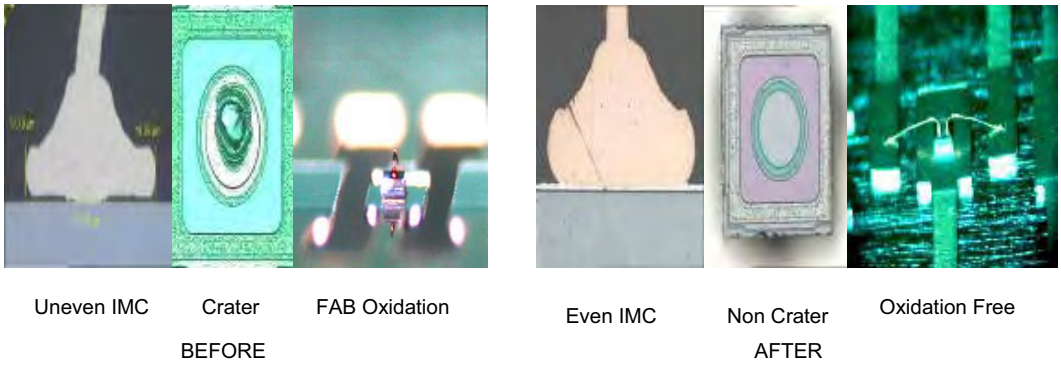


Photo 2. Performances comparison before & after changes.

Die Attach

Adhesion between die backside to bare Cu leadframe interface was studied, die shear strength was compared against control. Result met minimum shear requirement. Another concern was Cu leadframe oxidation, as eutectic bonding temperature >400°C was used. Leadframe staging at bond site evaluation showed oxidation when staging >60s.

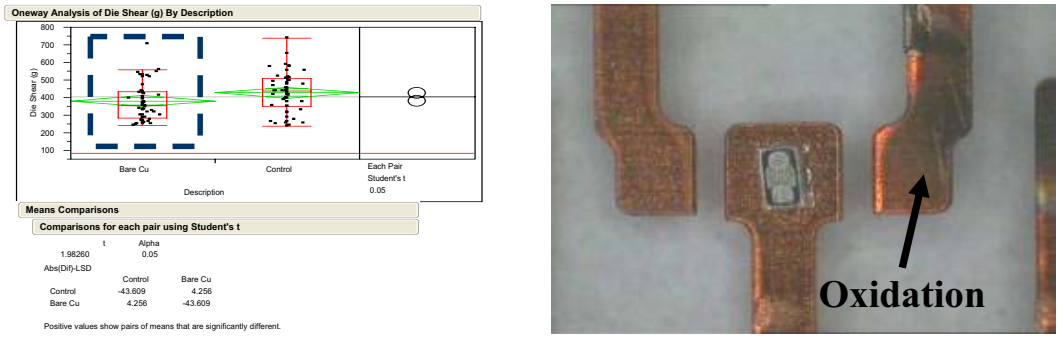


Figure 3. Die shear statistical comparison & leadframe staging oxidation.

Wire Bond

Typical bonding performances like ball shear, crater, wire pull, ball shape & stitch formation were assessed, using optimum Cu kit setup. Feasibility study showed acceptable results. Besides, evaluation on Cu leadframe oxidation impact against non stick on lead (NSOL) & breaking wire (BW) was carried out. Results revealed that NSOL & BW occurred at 40s leadframe staging at bonding area onwards.

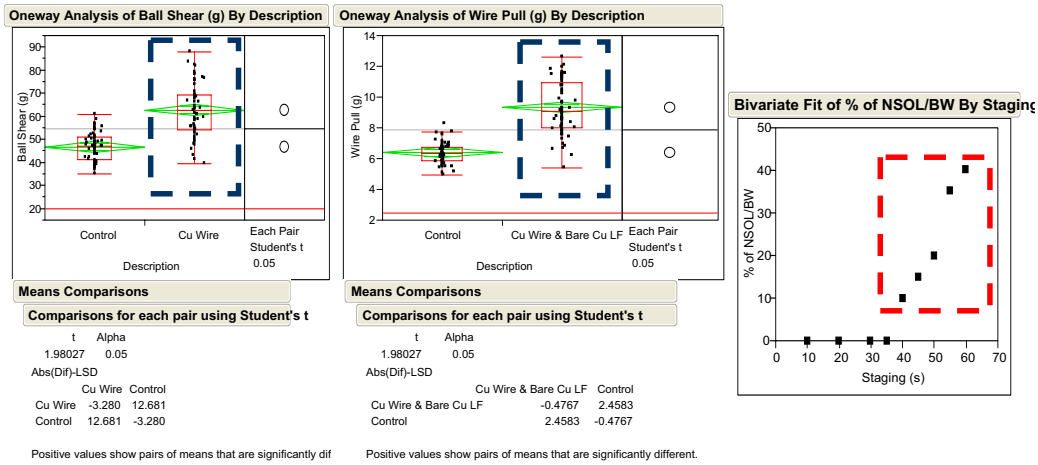


Figure 4. Ball shear & wire pull statistical comparison; NSOL/BW occurrences trend.

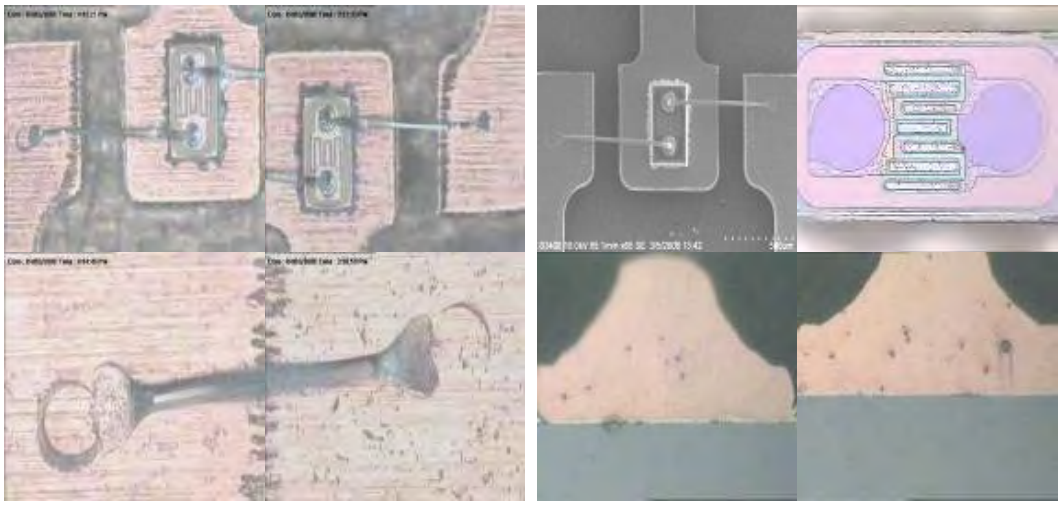


Photo 3. Cu^2 ball & stitch bondability assessment.

Molding & Plating

No moldability & delamination issues observed on bare Cu leadframe molding with current compound. At plating, solderability & plating thickness were monitored & passed requirements.



Photo 4. Bare Cu molded package outline.

Wire Bond Design Of Experiment

Design of experiment (DOE) on1st & 2nd bond was performed, to establish workable parameter setting for Cu[2] process. Table below is the test vehicle & DOE strategy.

TABLE I. Test Vehicle & DOE Strategy

Test Vehicle	Factors	Responses
Bonder : Shinkawa ACBxx Cu Wire : 22um, 99.999% Purity Capillary : Gaiser Leadframe : Bare Cu (non plated)	Full factorial design 1st Bond : Power, Force & Time 2nd Bond : Power, Force & Time	Ball Shear Cratering Non Stick On Pad (NSOP) Wire Pull Non Stick On Lead (NSOL)

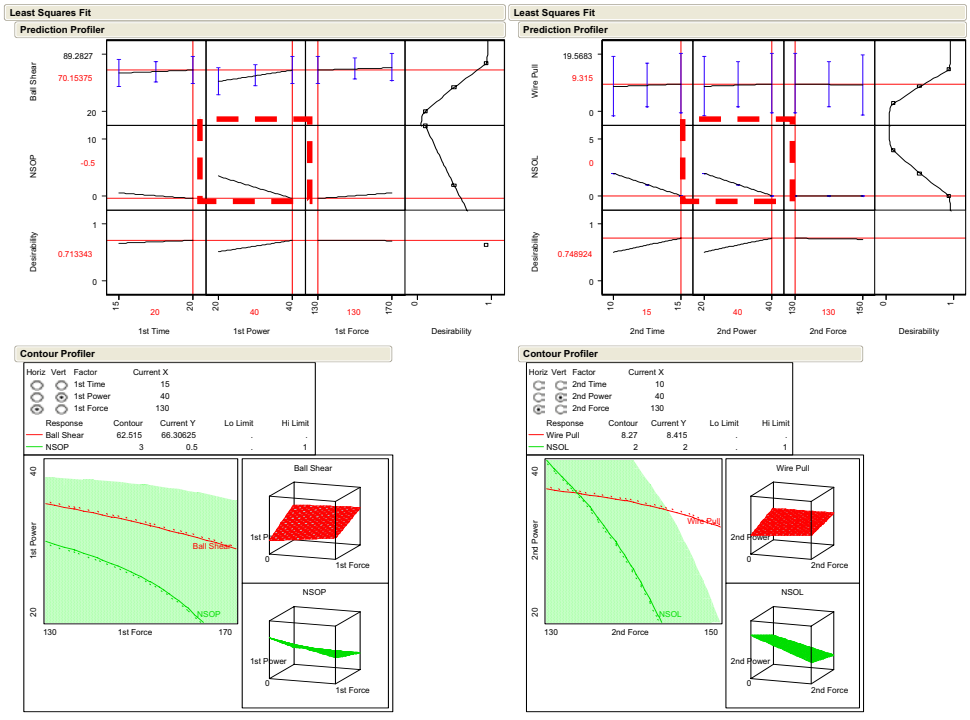


Figure 5. Prediction profiler & contour plot of 1st & 2nd bond DOE.

DOE results revealed that higher power would minimize NSOP & NSOL losses. All runs passed ball shear, wire pull & cratering requirements.

Pre-qualification Reliability

Units were submitted into in-house reliability test such as preconditioning, temperature cycle@1000x, pressure cooker test@96hours & high temperature storage@1000hours. No bonding related failure was observed on all the tests.

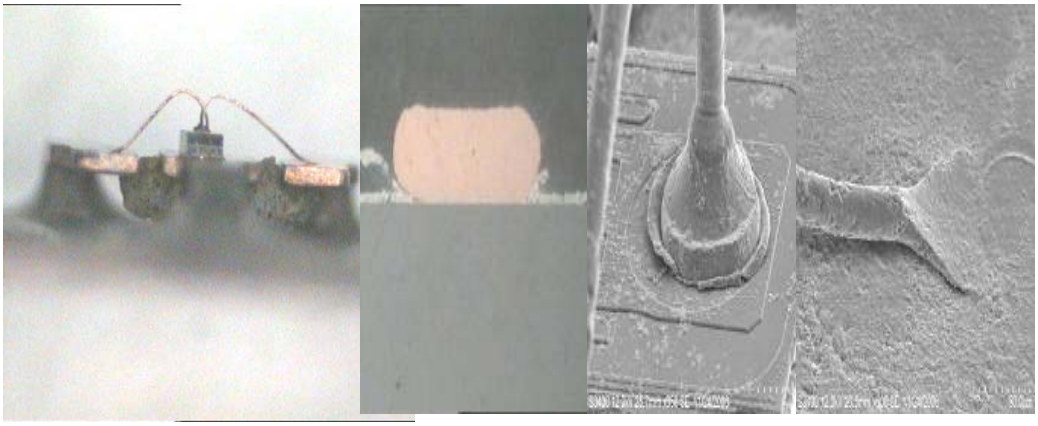


Photo 5. Post stress unit internal physical inspection.

Conclusion

Based on all the above studies, Cu^2 process could be manufactured for SOT package. However in order to have a robust process under mass production, Cu FAB & leadframe oxidation control shall be established & counter measure steps need to be in place. Besides, bonders' repeatability is also important to maintain consistent bonding. This process will be validated with more engineering samples build & official qualification.

Acknowledgments

The authors would like to thank Infineon Wuxi Mr. Chew Chee Khow (Managing Director) who is the site management driver; Mr. Neo Wee Min (Operation Head), Ms. Beh Paik Swan (Engineering Head) & Mr. Gooi Boon Huan on project management guidance and Mr. Zou Junfeng for reliability & failure analysis support. The authors also like to thank & appreciate Infineon Malacca Mr. Lim Chee Chian & Mr. Lee Cher Chia for their numerous supports on Cu wire bonding technical advice, Cu kit development & key information sharing.

References

1. Lim Chee Chian, Ng Kok Chai, Lee Cher Chia, Chai Min King, Lim Ong Seng, Chua Kok Yau, *Copper Wire Reliability And Bonding Integrity Robustness On Cratering Sensitive Bond Pad Structure,* IEEE External.
2. Ying Zheng, *Study Of Copper Applications And Effects Of Copper Oxidation In Microelectronic Package,* 2003.
3. Kulicke & Soffa (www.kns.com), *Bonding Wire Characteristics.*

ECS Transactions, 18 (1) 837-842 (2009)
10.1149/1.3096543 © The Electrochemical Society

Enabling Technology in Thin Wafer Dicing

Jeroen van Borkulo, Rogier Evertsen, Rene Hendriks,
ALSI, platinawerf 20G, 6641TL Beuningen Netherlands

Abstract

Driven by IC packaging and performance requirements, the die separation process is facing three main trends: Thinner wafers, stronger die and non-Silicon materials. Memory capacity increase, multichip functions and continuous package miniaturization require ultra thin wafer dicing. For IC performance increases, low-K material and other materials (top layers and even substrates) require different dicing methods for efficient and delamination free dicing processes. It is demonstrated that conventional blade dicing has reached the end of its capabilities and does not meet the separation requirements. Particular thin wafer development and the use of Die Attach Film (DAF) have added challenges and restrictions to the dicing technologies. Laser dicing processes have developed and have become the preferred dicing technology. This paper will discuss the various laser dicing technologies introduced for the different segments and will address the specific performances. Using a multibeam laser dicing process add significant advantages to the different solutions.

Background

Advanced Laser Separation International (ALSI) was founded in 2001 as a spin off from Philips Semiconductors. ALSI supplies semiconductor laser dicing systems using its proprietary multiple beam technology. ALSI is recognized by the semiconductor industry as a leading laser dicing technology company. ALSI is a market leader for laser dicing applications in RFIC and has a strong position and is a technology leader in the T&D and LED segment. Working with several “early adapter” semiconductor manufacturers to replace conventional mechanical wafer separation technologies, ALSI has achieved a large installed base and considerable experience in laser dicing and related infrastructures.

Challenges in thin Si applications

New IC technology, packaging as well as performance requirements drive the trend to thinner, more sensitive and complex structures and diversify the requirements in the die separation process. MEMS have fragile structures that do not allow dicing processes with high-pressure water cleaning, nor any dust or particle distribution.

With the further shrinking of IC dimensions, low-K material has been widely used to replace the traditional SiO interlayer dielectric (ILD) in order to reduce the interconnect delay. The introduction of low-K material onto silicon has imposed challenges on the dicing saw process. ILD and metal layers peeling and its penetration into the sealing ring of the die during dicing saw are the most common defects. Additionally, the blade costs went up and dicing productivity down for these applications.

The drive to more memory capacity per package volume, to enable the development of portable electronics, requires IC packaging with thinner wafers. 3-D packaging, stacked dies, often with Die Attach Film (DAF) are common and proven assembly technologies. Further wafer thickness reductions are needed but require new dicing methods to avoid yield loss at the stage where the wafer value is the highest or even worse when a broken die would yield out a package containing 10 or more dies.

The conventional saw dicing technology is preventing the ongoing trend for thin wafer development due to the following main reasons.

- Mechanical forces and vibration cause die crack and chip-outs

Ex. 1016, Page 865

- Short lifetime of saw blades due to inability to self-sharpening on the thin substrates
- Difficulty to cutting through the DAF with high dicing speeds
- Low assembly and die picking yield due to smearing of DAF against the die side wall.

As such major concerns and limitations for the application apply:

- Wafers are extremely sensitive to wafer stress (due to wafer thinning and front end processes)
- Dicing blade selection and exposure to material stacks in the dicing street (polymer, metal, passivation and other test structures)
- Dicing tape material, tackiness
- Dicing speed

Due to these issues and restrictions of conventional saw technology, the thin wafer IC and memory industry are searching for new dicing technologies and have identified laser dicing technology as the new separation technology. This new laser dicing technology has to comply with the demanding specifications.

1. A die strength value high enough to prevent die cracks during assembly and operating conditions
2. Compatibility to cutting through DAF, not influencing the DAF properties and resulting in 100% die pick yield
3. Ability to cut through a stack of different passivation materials and test structures (metals, polymers, oxides and nitrides)

Apart from these challenges, the laser dicing also offers additional opportunities:

1. reduction of dicing street width
2. High throughput (>12 wafers/hour)

Laser dicing technologies

Over recent years, several laser-dicing technologies have been developed each having their specific characteristics for a separation process. The main dicing technologies that have become most common are:

1. ***Ablation laser dicing***

   In this process the wafer material is removed by irradiation of laser pulses which locally generate a combination of melt and vapor. The vapor pressure drives the molten material out of the wafer generating an opening also referred to as kerf (see figure 1 below). This technology is predominantly used for dicing through the whole wafer substrate thickness, even backside metallization even though the use of backside metallization is not common for IC and memory applications. Depending on the application (wafer material, thickness, throughput, and die size), a certain laser type is chosen. Main laser process parameters that determine the interaction of laser light with the wafer material are wavelength, pulse duration and power. The size of the Heat Affected Zone (HAZ), recast and debris depend on these parameters.

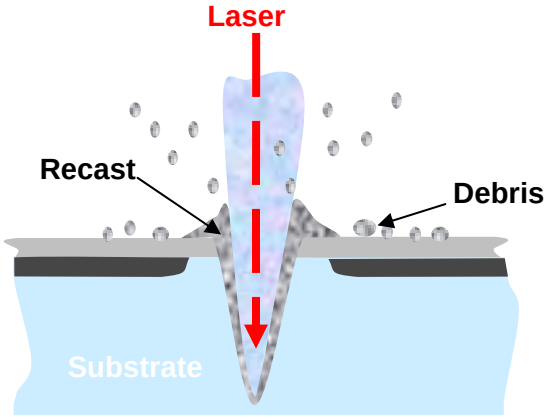


Figure 1. Removal principle for ablation

+ Ability to dice through top passivation, Si wafer and DAF
+ High removal rate therefore short dicing time

concerns with common laser technology:

– Reduces die strength
– Rougher side wall roughness
– Possibility of cracks and chip-outs when high power is used

2. *Subsurface dicing*

This separation process focuses laser pulses inside the wafer substrate generating a polycrystalline structure and therefore weakening the material locally (see figure 2 below). After processing the wafer can be expanded, sometimes with the aid of a breaking device, resulting in die singulation. This process works when there is no top passivation and test structures (specifically metal) within the street. These structures would block the laser radiation from going into the wafer material. In addition, any DAF present on the backside of the wafer is still not singulated and would need to be done in an additional step.

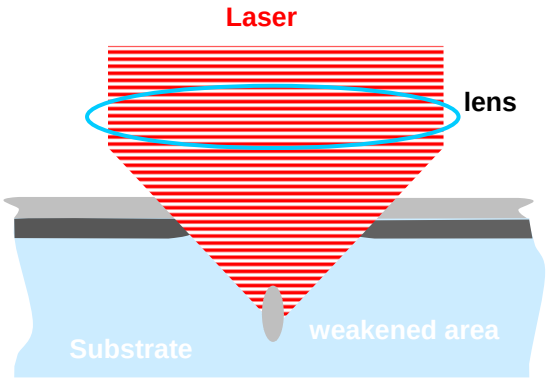


Figure 2. Subsurface principle

+ "zero" kerf width
+ smooth side wall surface

concerns with sub surface technologies:

– low productivity due to low process speed and alignment
– no DAF separation
– not compatible with structures in dicing street

3. *Hybrid laser technologies*

Several hybrid technologies combine laser and the conventional dicing saw or S&B. One typical example is the use of a low power laser to scribe through the top passivation and metal structures within the street, which the mechanical dicing saw has difficulties to cut through. In a subsequent step, the saw is used to cut through the actual Si substrate and possible DAF (see figure 3 below). This process has the advantage of a good interaction of the laser with the passivation and metal layers but still has the negative impact of the dicing saw to the die and the DAF tape (as mentioned above).

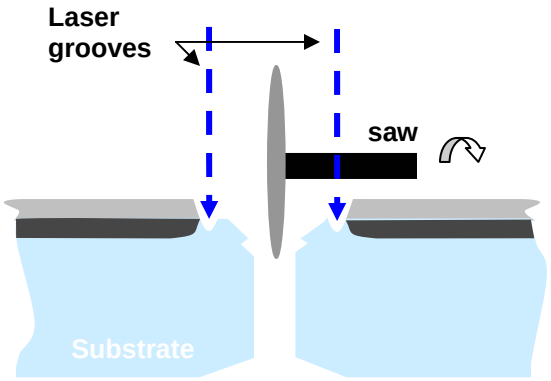


Figure 3. Hybrid principle combination of laser scribing and saw dicing

+ removal of passivation out of the dicing street
– typical sawing problems remain

Multiple beam technology

When using a laser dicing process to separate the wafer it is always a tradeoff between quality and speed, independent of the above mentioned technology. Nowadays, available industrial lasers can deliver high amounts of power. However, when exerted at such high levels to a thin wafer substrate, the material is not only separated, but may also be damaged severely. To get a good quality (no chipping or cracks) with a small or no HAZ, low laser

Ex. 1016, Page 867

power levels need to be used. As a result, material removal rates and therefore dicing speed is low and the laser capability with respect to the available power is far from utilized. To solve this issue Philips and ALSI have developed a proprietary multiple laser beam technology. The basic principle is to split the main laser beam up into a plurality of laser beams, each having a low power level and therefore not compromising the quality, but as a group of beams keeping the material removal rate and thus the dicing speed high.

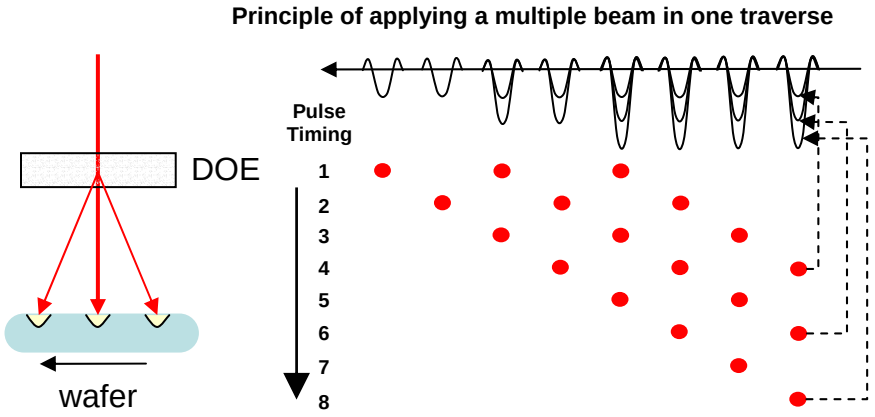
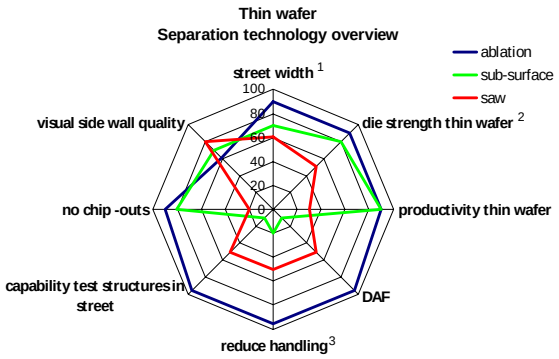


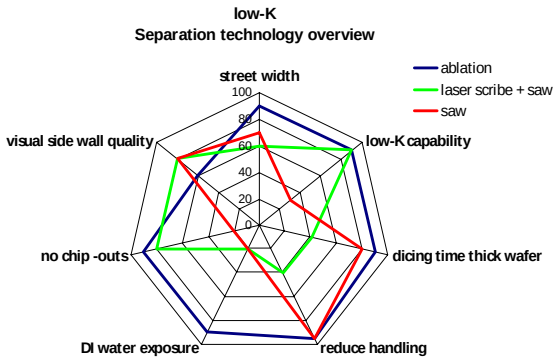
Figure 4. Multiple beam principle

In figure 4 indicated above the principle is demonstrated. The high power laser beam passes through a beam splitting device, in this example a diffractive optical element (DOE). This DOE splits the main laser beam up into a number of beams (in the situation illustrated above three beams) with a certain distance between them. The number of beams generated and the distance between them is depending on the design of the DOE; any number of beams and distance can be generated depending on the design of the DOE.

Using a multiple beam technology allows the dicing process engineer to tune to the optimal power levels for their application (ensuring minimum HAZ and high quality) and at the same time achieve a high throughput.

Using multiple beams gives the advantage of reducing number of traverses, whether either a laser ablation process is the best to use for an application or a subsurface dicing process. A multiple beam technology can be used for any laser dicing process and for each it brings the advantage of speed. When applying the multiple beam laser technology again significant advantages can be achieved and concerns for laser dicing are eliminated as demonstrated below in graph 1 and 2 for respectively this Si wafers and for low-K wafers.





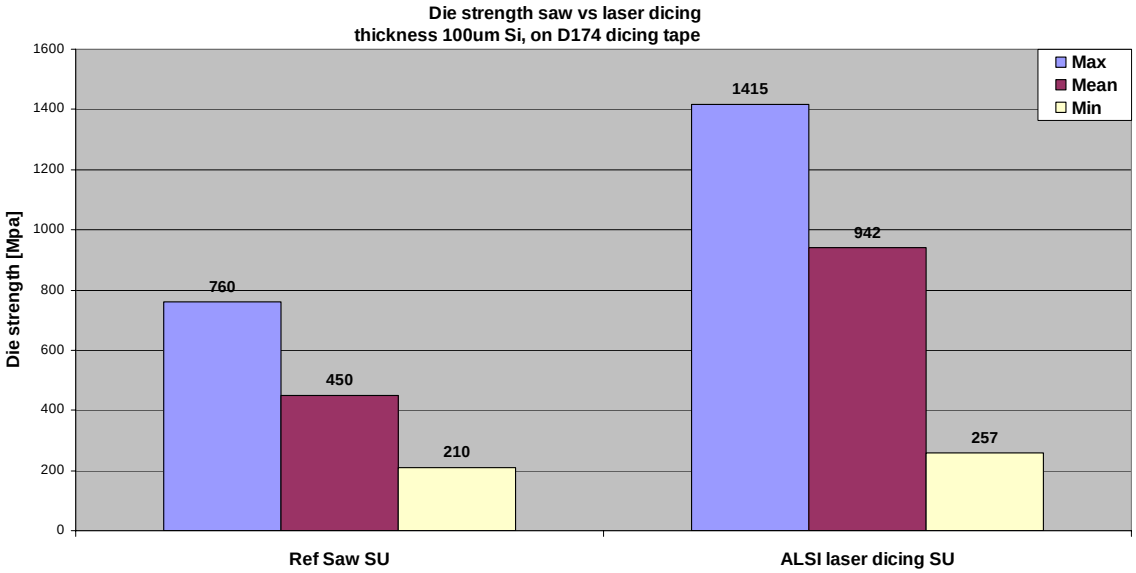
1) sub-surface; aspect ratio 0.4
2) Ablation; achieved via post process or short pulse laser
3) sub-surface; requires retaping, breaking, expansion

Graph 1. Thin wafer separation technology overview

Graph 2. Low-K separation technology overview

Post process

When wafers become thinner, they also become fragile and more likely to break when put under stress, specifically when the die strength value is influenced by the separation method. Therefore a new separation technology has to meet and even exceed the die strength values of the current conventional technology. When using a laser dicing technology to separate wafers, die strength can be lower compared to saw dicing technology due to the HAZ. This HAZ is determined mostly by the laser power and pulse duration of the laser irradiated at a certain location on the wafer surface. To minimize HAZ, laser power or pulse duration has to be reduced but as a drawback, the material removal rate is also reduced which means more traverses (more time) are needed to cut through the wafer material. Using a multiple beam technology already solves this problem of speed. If however, there is still too much HAZ and the die strength after laser dicing is too low, a post process can be used. In the post process step the recast material against the side wall of the die is removed and micro cracks that have been generated are made blunt such that they will not propagate further when put under stress. As the post process is fully integrated into the laser dicing system, each wafer coming out of the system has a die strength higher than the reference technology (see graph 3 below, SU = Structure Up).

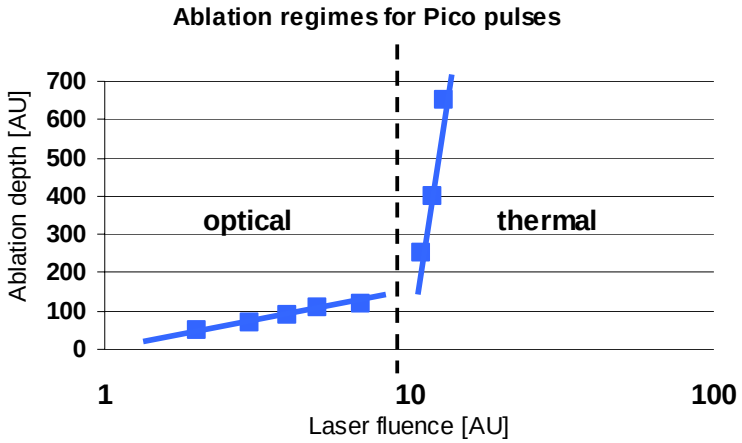


Graph 3. Die strength dicing saw vs. laser dicing

Future developments

To further improve laser material interaction and at the same time reduce the HAZ, it is beneficial to use even shorter laser pulses (pico second regime). When using laser pulses in this regime the laser material interaction is moving from a predominantly melt and vapor driven process (as for nano second pulses) to a vapor and bond breaking process (also called optical or ionization process). In a pico second laser pulse, all energy is supplied to the wafer surface within an even shorter period generating extremely high fluence. These high power densities have the capability to pull all the electrons from the atom structure and therefore breaking the bond structure between them. This is also referred to as “cold dicing” but is only valid when relative low pulse energies are used. This obviously also means that the amount of material removed per pulse is low. When the power (fluence) of pico second pulses is increased, the process is changing from a predominantly bond breaking process to a thermal process equal to nano second pulse lasers and basically the advantage is lost (see figure 7 below).

Ex. 1016, Page 869



Graph 4. Ablation regimes for Pico second pulses

The transition to a thermal process occurs because the plasma generated by the ionization in the first stage of the pulse is becoming too hot and is starting to “dice” into the wafer material on its own, generating melt and more HAZ. Therefore, it is necessary in order to use the advantages of pico second pulse laser to keep the pulse energy per beam low. When this happens the HAZ is minimized and the recast layer against the sidewall is also of a different structure. The recast layer for nano second pulses consists of molten Si which did not get ejected out of the kerf. For a pico second laser the recast is merely a thin oxide layer, which also acts as a passivation layer on the side wall and aids in the die strength value. The first die strength tests demonstrate values close to the saw reference and there is ongoing development being done (more details will be presented during the oral presentation). As low powers need to be used to utilize the advantages of pico second pulses, the initial dicing speed will also be low. However, the multiple beam technology also brings a good solution in this application, for a pico second laser. In this case high power levels are available allowing splitting up the pico second laser pulse into a plurality of beams ensuring a high total removal rate and speeds, higher then currently possible with the conventional dicing saw.

Conclusion

Memory and IC manufacturers are facing difficult challenges to keep up with the strong demand for packaging. Substrate thicknesses of 50um and less on DAF tape force them to look to new separation technologies. Laser dicing is providing them a solution to meet the strict specifications on the diced die. However, no matter what laser dicing solution is chosen, it will always be a trade off between quality and speed. The multiple beam technology from ALSI provides a solution to go around this impasse and maintain both high quality and speed. Each of the technologies supplied by ALSI; ablation (nano or pico) or subsurface, is equipped with this multiple beam technology to ensure better dicing quality and higher yield, in combination with a quick return of investment, in laser dicing technology.

CHAPTER 8

EMERGING SEMICONDUCTOR TECHNOLOGIES

ECS Transactions, 18 (1) 845-850 (2009)
10.1149/1.3096544 © The Electrochemical Society

Growth and Electrical Characterization of Horizontally Aligned CNTs

C. Santini[a, b], D. J. Cott[a], A. Romo Negreira[a, c], S. Riva Sanseverino[b], S. De Gendt[a, d] and P. M. Vereecken[a]

[a)] IMEC, Kapeldreef 75, 3001 Leuven, Belgium
[b)] Department of Electrical, Electronics and Telecommunication Engineering, University of Palermo, Viale delle Scienze, 90128 Palermo, Italy
[c)] Metallurgy and Materials Engineering Department, KU Leuven, Kasteelpark Arenberg 44, Leuven, B-3001, Belgium
[d)] Chemistry Department, KU Leuven, Celestijnenlaan 200f- Leuven, B-3001, Belgium

Nickel catalyst nanoparticles acting as nucleating seeds for carbon nanotube (CNT) growth were selectively deposited on the sidewalls of titanium nitride (TiN) electrodes by electrochemical deposition (ECD) from a Ni^{2+} electrolyte solution. Horizontal aligned CNTs were grown selectively from the sidewalls of these TiN electrodes forming a bridging CNT contact between the electrode gaps. Current-voltage measurements demonstrated this could be a promising technique towards nanoscale interconnections and nanoelectronic devices with resistance values for the bridging CNTs as low as 300Ω.

To realise the potential of CNTs in microelectronic applications (1), sufficient control over their location and orientation should be demonstrated. Typically, control over the horizontal growth of CNTs has been achieved using controlled gas flows (2), electric fields (3) and CNT-surface interactions on faceted substrates (4). Herein, we pursue an alternative approach, by placing the CNT catalyst selectively on the horizontal sidewalls of prefabricated TiN electrodes. CNTs were grown across TiN electrodes of 200 nm thickness fabricated on 200 nm oxidised p+ implanted doped silicon substrates. A cross-sectional view of these electrodes is outlined in Figure 1a. There is a direct connection between one (Figure 1a) or both electrodes through nickel silicide pads and the Si substrate. For clarity we will denote both electrode configurations as A and B, respectively. In configuration A only one electrode is connected to the substrate whereas configuration B has both connected. One of the electrodes in B can be disconnected after catalyst placement and CNT growth (*vida infra*) to avoid a competing electrical pathway through the Si substrate. On the top surface of the electrodes an insulating 30 nm thick silicon oxynitride (SiON) hard mask was deposited. These electrode structures are connected within a 4-point probe configuration (Figure 1b). This allows for accurate CNT resistance measurements by eliminating contact resistances and the resistances of the leads.

The first challenge in achieving CNTs bridging the electrodes is to deposit metal catalyst on the sidewall of the electrodes. This can be routinely achieved using such methods as physical vapor deposition (PVD) , spincasting (5), and printing (6). However, all these methods lack selectivity and require further steps to isolate the catalyst at specific locations. Electrochemical deposition (ECD) has an intrinsic selective nature in that metal deposition from solution only occurs on conductive surfaces contacted as the

working electrode. Therefore, this method has the advantage of requiring no further processing steps to remove or clean unwanted metal catalyst deposits.

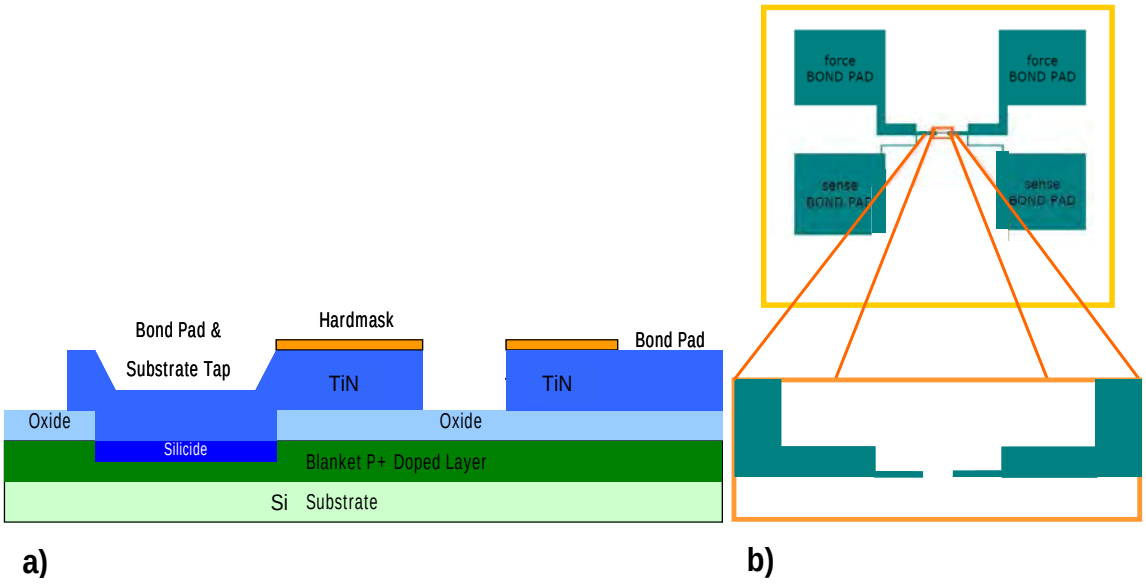


Figure 1. Schematic of CNT growth test structure: (a) cross-section view showing the different layers and one electrode connected to the doped Si substrate through the oxide. (b) top view of the probe pads for four-point probe tests and a magnified view of the individual electrode gap.

Electrochemical deposition of nickel nanoparticles was carried out from a Ni^{2+} electrolyte solution employing a similar procedure reported previously (7, 8, 9). Briefly, an electrolyte solution was made up of 0.01M $Ni(NO_3)_2$, 1M NaCl, and 1M NH_4Cl. The pH of the solution was adjusted to 8 by adding small volumes of NH_4OH. Prior to deposition, the samples were pretreated in NH_4OH solution (2.9 wt. %, 3 minutes) to remove any polymer residues followed by immersion in dilute HF (0.25 wt. %, 1 minute) to remove the native surface oxides. Electrical contact was made to the backside of the Si substrate using InGa eutectic. The electrode potential was controlled with respect to a Ag/AgCl/3M NaCl electrode (0.21V vs. SHE) and a Pt wire was used as a current collector or counter electrode.

Figure 2 provides a clear demonstration of the selective nature of the ECD technique, comparing the electrode-substrate contact configurations A (Fig. 2a) and B (Fig. 2b). Catalyst deposition occurs only on those TiN electrodes that have a direct contact through the insulating bottom SiO_2 layer to the Si substrate. The Ni nanoparticles nucleated only on the side walls as the SiON hardmask layer on the top of the TiN surface presents a barrier to current flow. Through variation of the deposition time and applied potential it is possible to control the size and surface density of the nickel nanoparticles. In particular, by increasing the deposition time from 10 to 100 ms the average diameter increases from 14 to 30 nm at -1.3V. In addition, the estimated (SEM) nucleation density varies from 10^{10} to 10^{11} NP/cm^2 when changing the applied voltage between -1.3 and -1.6V for a constant deposition period of 50 ms.

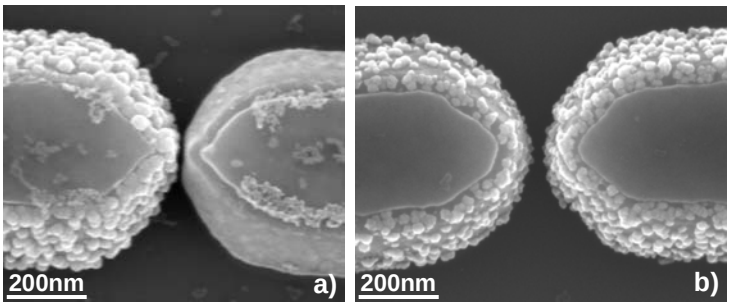


Figure 2. SEM images of nickel nanoparticles electrodeposited on the sidewalls of TiN electrodes with (a) only the left electrode contacted to the Si substrate and (b) both electrodes connected to the substrate.

Next CNTs were grown by chemical vapor deposition (CVD) (10) from the electrodeposited Ni catalyst nanoparticles (ECD voltage: -1.8V, ECD time: 0.1s). By varying the growth parameters an effective CNT growth process was determined which produced high-density multi-walled (MW) CNTs on the TiN sidewalls. Samples were exposed to 10/150/600 sccm of a mixed gas flow of $C_2H_2/H_2/Ar$ at 4 Torr (200mm Oxford Instruments Nano PECVD reactor). The substrate temperature was maintained between 575-600 °C for periods up to 30 minutes. Figure 3 shows an example of TiN electrodes after growth for the two different electrode-substrate configurations A (Fig. 3a) and B (Fig 3b).

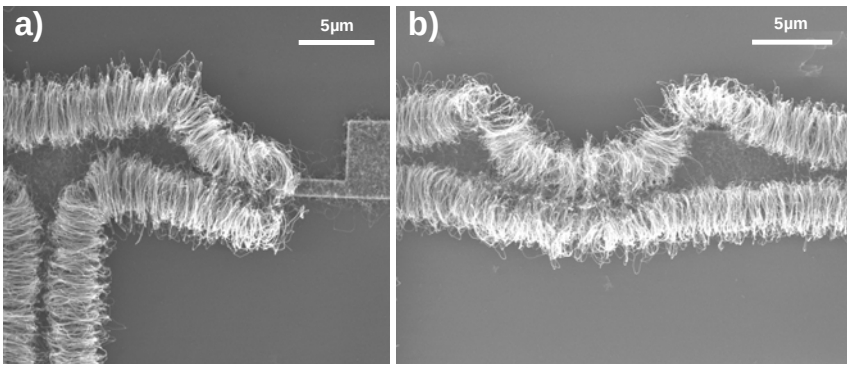


Figure 3. SEM images of TiN electrodes after CNT growth. with a) one electrode connected to the Si substrate b) two electrode connections.

With Ni catalyst only on one of the electrodes, CNTs grew from one electrode towards the other in Fig. 3a. With Ni catalyst on both electrodes, the CNTs grew simultaneously from both electrodes in Fig. 3b. The CNT bridging the gap seem entangled even though some may form direct contact with the opposite electrode tip.

A good correlation was found between the average size of the nanoparticles and the CNT growth: e.g. 15-35 nm Ni nanoparticles produced CNTs of 10-25 nm average diameters. The small difference in nanoparticles and CNT diameter is most probably attributed to the reshaping of the Ni particle during growth (spherical versus elongated). CNTs were over 5 µm in length and curved as they initially grew directly out from the sidewall in a direction parallel to the substrate surface, bended backwards and continued growing in the opposite direction above the top surface of the electrode. In fact, for a

growth time of 5 minutes only, the CNTs grew for about 800 nm perpendicular from the electrode sidewall without bending backwards (data not shown). Therefore, after a certain CNT length, a condition arises where the CNT bundles are attracted back over the top surface of the electrode. It is well documented, that tube to tube Van der Waals interactions within dense CNT forests create contracting forces towards its centre thus maintaining CNT alignment. However, a recent report (11) argued that Van der Waals interactions between CNTs and an adjacent structure were sufficiently strong, enough to alter the growing direction of CNTs. To date, we cannot discount other contributing factors from the local gas environment.

4 point probe electrical measurements were carried out for both electrode-substrate configurations, i.e. for CNTs grown from one electrode to the opposite (configuration A) and CNTs grown from both electrodes simultaneously after cutting the connection to the substrate of one of them (configuration B). Figure 4 shows I-V measurements for CNTs bridging a 400nm gap between the TiN electrodes with CNTs grown from both electrodes as shown in the inset of Figure 4b. Average CNT diameter was around 25 nm. As the tubes are multiwalled, they are metallic at room temperature (12) and Ohmic behavior was indeed observed at lower voltages. When increasing the voltage up to approximately 1.5V the resistance value decreased from an initial value of 880Ω to 400Ω (Fig. 4b). At higher applied voltages, the current dropped and the resistance increased significantly. A certain proportion of the CNTs seem to break, most probably 'burned' due to the Joule effect. It has indeed been observed (13) that the electrical breakdown of CNTs can be induced by non-homogeneously Joule heating along a CNT without release of heat into the contacts as CNTs broke down at the mid-length of a 1-2 μm gap.

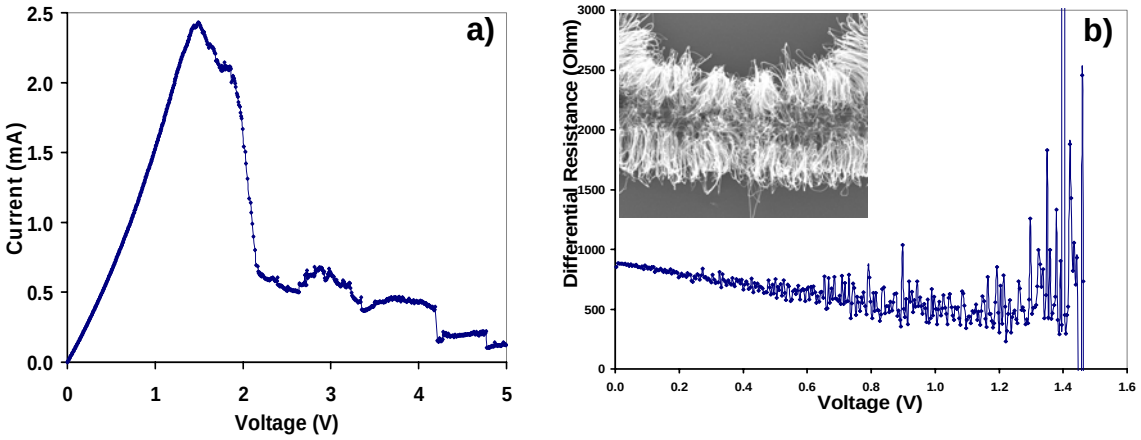


Figure 4. 4-point I-V measurements carried out on 400nm TiN electrode gap with CNTs grown from both electrodes: a) Current vs. Voltage, b) Differential Resistance vs. Voltage until the CNT burn. Inset of Figure 4b: SEM image of the CNT bridges before electrical measurements. It is difficult to identify rupturing of individual CNTs after these measurements due to their high densities.

Similar trends were found for the various samples fabricated with different catalyst deposition conditions and/or CNT growth time (length). Outlined below in Table I; increasing the gap spacing increased the measured resistance, and for CNTs grown from both electrodes (configuration B), lower resistance values were obtained than for CNTs bridging from one electrode only (configuration A). Not only is the density of CNTs

Ex. 1016, Page 876

approximately double (and hence resistance half), but CNTs grown from opposite sites entangle and cause shorted paths for current to flow even when they do not directly bridge the gap.

TABLE I. Resistance measurements (kΩ) for CNT bridging TiN electrodes for different effective gap spacings and configurations.

Configuration electrode-substrate	Gap spacing	Average differential resistance values (kΩ)
Configuration A	200nm	2.15
	400nm	7.5
	600nm	12
	1µm	17.5
	2µm	14
	5µm	112
Configuration B	200nm	0.14
	400nm	1.4
	600nm	1.55
	1µm	2.45
	2µm	2.2
	5µm	17.5

Figure 5 was obtained by scanning the current through a CNT bridge (configuration B). It can be seen that a sudden drop in voltage at 0.18mA indicates that new contacts may have formed by electrostatic trapping. Indeed, the second and third scans show a reduced resistance value. Owing to the high density of CNTs between the electrodes (figure 4b inset) this enhanced current transport can arise from a number of scenarios. There will be a lack of direct electron pathways across overlapping CNTs and this may actually facilitate current-induced welding at points where CNTs overlap (14). In fact, high current created hot spots can promote cross linking of dangling bonds at CNT defect sites (15) which enhance electrical pathways and could lower the overall resistance of the CNT bridged electrodes.

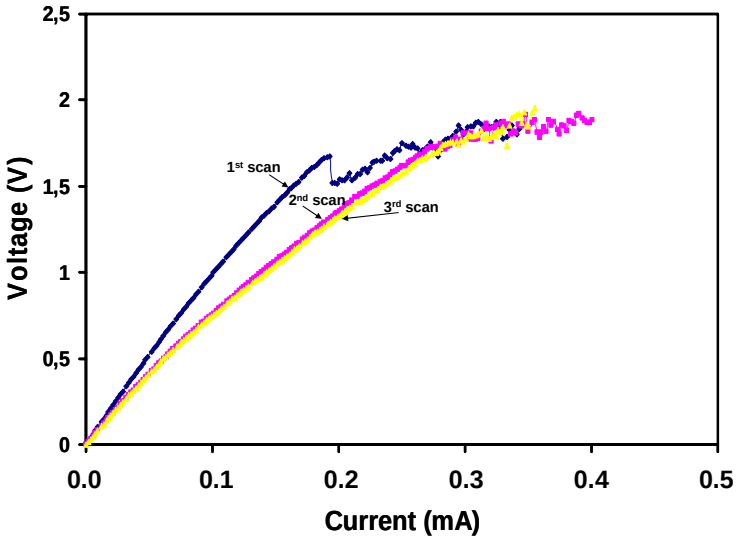


Figure 5. Successive I-V sweeps carried out on 1µm, configuration B gap spacing TiN structure below the burning condition.

In conclusion, the results outlined demonstrate that electrochemical deposition (ECD) is an effective method to selectively deposit metallic nanoparticles (Ni) on non-planar conducting surfaces. For this to be realised, a direct contact or electrical pathway between

Ex. 1016, Page 877

the area of deposition and the working electrode (Si substrate) must be present. This method offers excellent control over density (~10^{11} NP/cm^2) and size (15-30nm) of the catalyst particles at room temperature and is directly translated to the density (~10^{11} CNT/cm^2) and diameters (10-25 nm) of the CNTs. 4-point probe electrical measurements demonstrated that a contact between two electrodes is created by the CNT bridging. This occurs when CNTs grow from one electrode to an opposite facing electrode and also when they grow simultaneously from two opposite electrodes. This last configuration allows for lower resistant contacts than the other one simply due to the fact there should twice the number density of as grown CNTs. Voltage drops during current scans indicate an improvement in the contact due to electrostatic trapping.

Acknowledgments

The authors would like to acknowledge Bernard Capraro, Gavin D'arcy, Paolo Rapposelli and Ross Kennedy from Intel Ireland for design and fabrication of the electrode test structures. This project was funded within the CARBonCHIP project supported by the European Commission, contract NMP4-CT-2006-016475 and IMEC's Industrial Affiliation Program. C. Santini acknowledges the Marie Curie EST project 'APROTHIN' that allowed her to carry out this research at IMEC, Leuven, Belgium as part of her PhD program.

References

1. M. J. O'Connell, *'Carbon Nanotubes: Properties and Applications'*, 2006
2. S. M. Huang, X. Y. Cai, J. J. Liu, *Am. Chem. Soc.*, **125**, 5636 (2003).
3. A. Ural, Y. M. Li, H. Dai, *J. Appl. Phys. Lett.*, **81**, 3464 (2002)
4. C. Kocabas, S. H. Hur, A. Gaur, M. A. Meitl, M. Shim, J. A. Rogers, *Small*, **1**, 1110 (2005)
5. N.R. Franklin and H. Dai, *Adv. Mater.*, **12**, 890 (2000)
6. Z. R. Abrams, Z. Ioffe, A. Tsukernik, O. Cheshnovsky, Y. Hanein, *Nano Lett.*, **7**, 2666 (2007)
7. A. Romo Negreira, P. M. Vereecken, C. M. Whelan, G. Groeseneken and K. Maex, *ECS Transactions*, **2**, 409, (2007)
8. A. Romo-Negreira, D. J. Cott, A. S. Verhulst, S. Esconjauregui, N. Chiodarelli, J. Ek-Weis, C. M. Whelan, G. Groeseneken, M. M. Heyns, S. De Gendt, P. M. Vereecken, *MRS spring proceedings*, 2008, in press
9. A. Romo-Negreira, O. Richard, S. De Gendt, K. Maex, M.M. Heyns and P.M. Vereecken. *Sci. Adv. Mater.*, **1**, 1, (2009), accepted.
10. D. M. Dobkin, M. K. Zuraw, *'Principles of chemical vapor deposition'* 2003
11. A. Kawabata, S. Sato, H. Shioya, T. Iwai, M. Nihei, D. Kondo, and Y. Awano, *Japan. J. Appl. Phys.*, **47**, 1975 (2008)
12. M. Wang, J. Wang, Q. Chen, L-M. Peng, *Adv. Func. Mat.*, **15**, 1825 (2005)
13. Y. Lee, J. H. Lee, S. J. Chung and S. Lee. *App. Phys. Lett.*, **89**, 073109 (2006)
14. M. Stadermann, *Phys. Rev. B*, **69**, 201402(R) (2004)
15. S. Agrawal, M. S. Raghuveer, R. Kröger, G. Ramanath, *Journal of Applied Physics*, **100**, 094314 (2006)

ECS Transactions, 18 (1) 851-856 (2009)
10.1149/1.3096545 © The Electrochemical Society

Fine Pitch Transfer and Join (T&J) From MCM to 3DI

R. R. Yu

IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598
yur@us.ibm.com

Over the last three decades computational performance propelled forward at a breath taking pace. This was enabled due to the introduction of CMOS technology, strict application of its scaling rules, and system innovation. The industry followed "Moore's Law" doubling transistor count every 18 month. In addition the design community drove innovations like on-chip caching, pipelined executions, and multi-core systems to deliver the required performance. All this time the processor chip could be considered to be flat which put increasing pressure on chip size to accommodate all its components. To continue the performance trend, however, it is conceivable to spread the components of the microprocessor chip over more than one active layer to maintain their close proximity. The device 3D integration (3DI) has the potential to achieve such goals to the next level. 3DI can be visualized as vertically stacked multi-chip-modules (MCM-D) with through-Si-via (TSV) integrated with all Si devices. The fine pitch transfer join (TJ) technology is well developed and suited for such integrations. This paper will review the evolution of the TJ technology in fine pitch device connection from it's early MCM applications to the more recent all-Si fine pitch 300 mm wafer-level connections.

Introduction

In the next decade we will see the development of computing systems fall into the two categories: "More Moore" and "More than Moore". "More Moore" will continue the scaling trend, however in various modifications. The main part will still be geometrical scaling to achieve the highest possible integration density for cost reasons. There will also a design equivalent scaling that is mostly driven by the goal to achieve the best power performance trade-off. The recent introduction of the multi-core systems is an example for this type of scaling: An assembly of a number of cores with lower frequency is able to execute the same number of operations per second than one high frequency core. The challenge in this scenario is, however, to make sure applications can take advantage of a high degree of parallel processing. The problem of performance gain is now given to the programmers and soft-ware architects. In this multi-core architecture we will see a tremendous stress, for some applications, on on-chip cache capacity that in turn will put pressure on the chip size. A way out is to stack a memory layer on top of the multi-core layer, for instance. This would allow maximizing on-chip cache (close proximity) and large cache capacity. Both layers would benefit from the geometrical scaling.

"More than Moore" will address a different scenario which is not so much related to pure performance but more to functional diversification or modularity of the system. The value preposition for the customer is the combination of more advanced functions in the

Ex. 1016, Page 879

system or a high flexibility in its assembly. 3D integration can be used to combine incompatible technologies like: Si-CMOS and sensors, Si-CMOS and optical, Si base with III/V base. High performance with low cost, etc. Many other combinations are possible.

For both "More Moore" and "More than Moore" 3D integration can serve a valuable technology to boost performance or versatility of the computing system.

.

This paper will discuss 3D integration with the transfer-join (TJ) method. This TJ method for 3D integration was initially developed for IBM high end system thin film multi-chip-module (MCM-D). The technology was then further enhanced into a wafer level bonding, thinning, and fine pitch transfer joining connection technology with and/or without the aide of a glass carrier. We will illustrate the versatile nature of this technology in wafer level applications for fine pitch CMOS integration, in wafer thinning, in MEMS application and recently in extending to 300 mm wafer level 3DI connections.

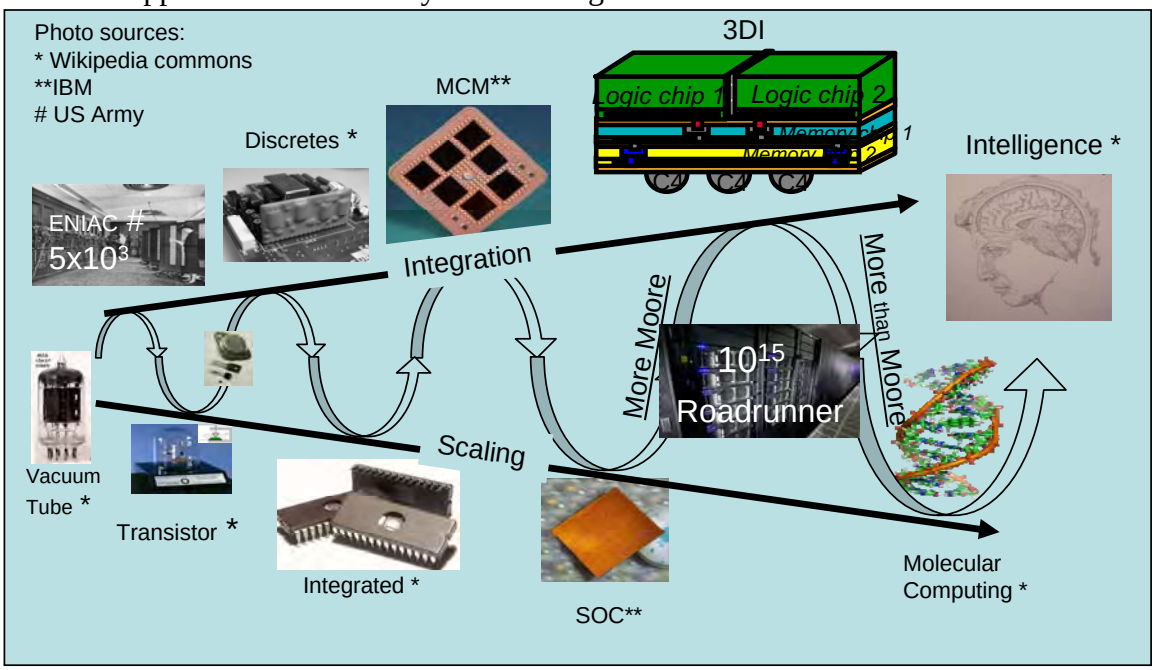


Figure 1: Two governing principles: integration (Amdahl's law) and scaling (Moore's Law) for computing system performance.

Transfer join (TJ) from MCM to 3D

3DI introduces a new integration dimension in devices and various methods have been proposed to address the associated challenges (8). The approaches can be grouped into a) package-centric (chip-chip, or C2C), b) BEOL-centric (wafer-wafer, or W2W), and c) hybrid die-wafer (D2W) options. In the package-centric approach the emphasis is on known good die (KGD) and on chip-chip or passive-chip integration where the wiring density improvement is secondary to functionality. Multi-stacking and lower cost are possible but with relatively higher difficulty (8, 9). In the BEOL-centric approach the focus is on wafer level connections and potentially with a high TSV density and lower cost (8, 10). The D2W approach is to populate the KGD on a wafer like D2D and process the assembly as a pseudo-wafer like W2W, resulting in benefits from both D2D and

Ex. 1016, Page 880

W2W (11). A detailed analysis of various approaches and their advantages can be found in (8, 12).

In this paper we will focus only on the original Cu/adhesive hybrid connection method and its various applications in wafer-to-wafer and die-to-wafer connections.

1) MCM/SCM Thin film Transfer and join (TFTJ):

The wafer level transfer and join (TJ) technology originated from MCM-D thin film transfer and joining (TFTJ) technology developed for an IBM high-end systems. In order to enhance the MCM-ceramic (MCM-C) wiring density, a low-k polyimide ILD thin film (TF) wiring plane pair was added to the package. This TF package was originally added to the MCM-C base through a serial build process. A parallel build for the TF layers on glass plate as carrier was later developed (13). The TF can be built on the glass carrier with either a face-up or a face-down approach (12). The face-up approach is for a complex system with full level test and repair capability from both top and bottom surfaces of the TF modules (12). To better manage the film distortion during the build a two-step transfer method was developed. First the face-up TF (top surface up) was built, tested and is bonded to a temporary glass carrier using a Teflon-PFA adhesion layer. The connecting IOs (bottom surface) are then prepared after the flip and tested for the bonding interface and the TF is then transferred from the temporary glass carrier to the permanent MCM-C base through the bottom surface IO. The Teflon-based adhesion layer on the top surface also serves as a compliant layer during the lamination to the base MCM-C carrier. This modular "flip" approach (starts from top-up and ends on bottom-up on a temporary carrier) ensures the TF always attaches to a rigid carrier and is the first successful film transfer method on a modular level (Figure 2b from (12)). This original flip process has been widely adopted in various "face-to-back" chip and wafer stacking methods (9, 14).

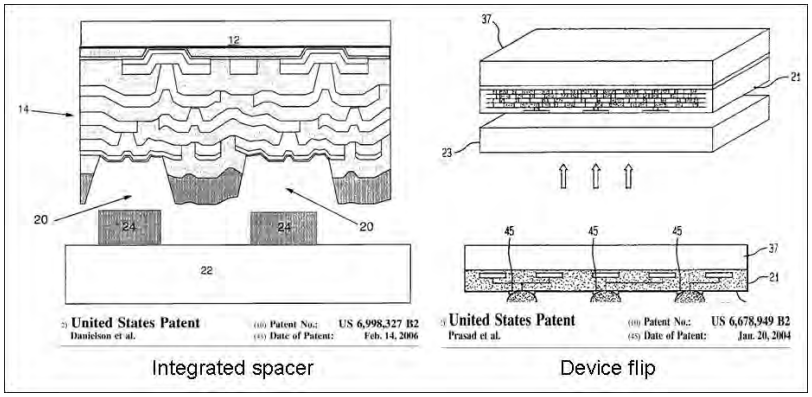


Figure 2: a) Integrated spacer (12), b) Device flip for face-back connection (12).

For the simpler single chip modules (SCM) the TF is preferred to build face-down (structure reversely built with bottom-surface up) and the TF is transferred directly from the glass carrier to the MCM-C base in one transfer step (Figure 2a, also from (12)).

In the final transfer bonding of the TF to its MCM-C base, both integrated and discrete spacers can be used as the bonding structure. Discrete spacer is a separate bondable material with the through-via punched in for IO connections. Figure 3a shows an MCM-D module built with discrete spacer with the typical SEM cross-section of the spacer structure. Figure 3b shows several SCM modules with integrated spacer. In both cases

Ex. 1016, Page 881

polyimide adhesives were used to bond the layers together. 90/10 PbSn solder was used for metallurgical connections (12).

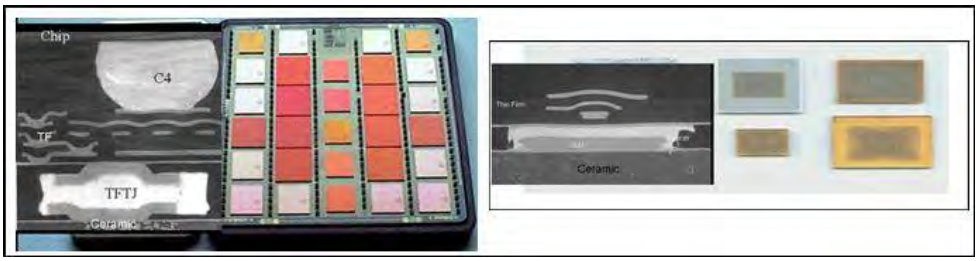


Figure 3: a) MCM-D with discrete spacer (12), b) SCM with integrated spacer (12)

3) 3DI Wafer-level bonding and thinning

The bonding and transferring of TF module using glass carrier formed the basis for wafer level 3DI bonding, thinning, transferring, and joining within IBM. The detailed process flow can be found in (12). Figure 4a shows an 8" device wafer initially bonded to a carrier glass with the adhesive, was then thinned to 5 μm, and is released from the carrier as a stand-alone wafer membrane. Figure 4b shows that the bonded wafer can be polished to remove all the bulk Si to its buried oxide (BOX) surface with devices left only (. Figure 4c and 4d show 2000A Si structure were thinned and transferred to a CMOS base. Both Teflon-based and polyimide-based adhesives were evaluated for use in lamination at and beyond 300C (11, 12, 15, 16).



Figure 4, a) 5 um Si film (12), b) Si etched to BOX (14), c) and d) 2000A Si transferred to a CMOS driver chip (16)

4) Through-Si-via (TSV)

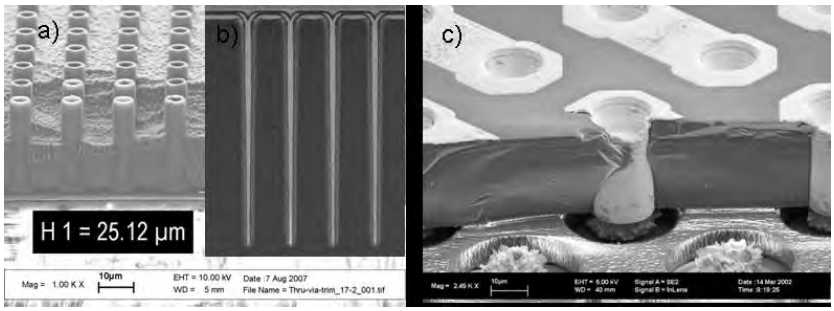


Figure 5: TSV: a) 5 x 25 um (17) W, b) 1.5 x 27 un W (17), c) 15 x 20 um Ni (12)

With the wafer bonding and thinning to sub-20 μm capability, TSV process with high density connection becomes possible. Several ways of deep Si via fill have been

evaluated. The initial work showed that a 15 μm critical dimension (CD) 45 μm pitch at 20 um deep TSV with electroless Ni was possible (16). More recently the TSV metal is more focused on tungsten (W) for deep via fill (9, 17). W has a close TCE with Si and the CVD process is more suited for deep via fill.

5) W2W and C2W interface via connection (IVC) and post thinning BEOL:

The integrated spacer concept developed for the MCM/SCM transfer bonding has been advanced to a finer pitch and was extended to wafer level transfer and joining (TJ) technology (11, 12, 15, 16, 17). Figure 6c shows the original alignment and bonding configuration for the TJ connection (11). Figure 6a are the 10 um lines at 20 um pitch via chains after the transfer bonding while figure 6b shows one of such x-section (11). The overlay accuracy by using the stud/via lock/key structure on a wafer level has been demonstrated to be within 1 μm. The via-recess is normally a double layer structure with the lower layer as a distortion restrainer and the top layer an adhesive. The via/stud size has be demonstrated ranging from 1 μm to 20 μm with a typical 5 μm CD and a height of 3 to 5 μm in most cases (11, 12, 17).

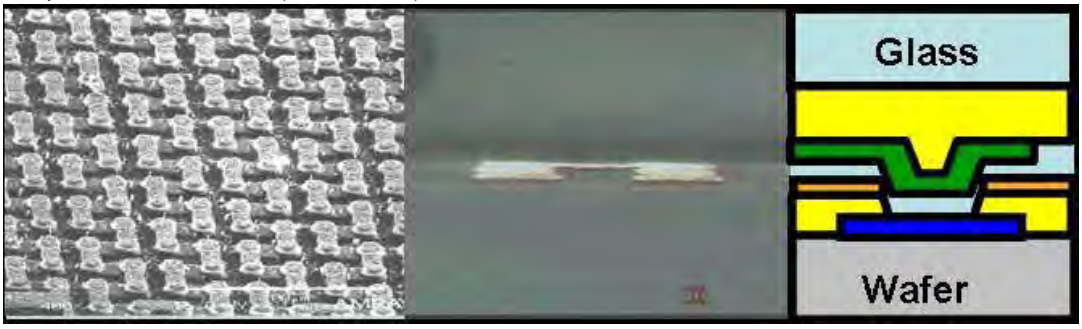


Figure 6: 1) 10 x 20 um via-chain after transfer, b) x-section, c) Bonding structure. All from (11, 12)

Recently, the TJ Cu/adhesive hybrid bonding has been extended to 300 mm wafer format (17). The complete via chains including TSV, IVC, and BEOL had been well yielded for various via pitches at Si thickness around 20 um. Figure 7a shows one of such completed 300 mm 3Di wafers. Figure 7b shows the SEM x-section of the 3DI structure with the view of TSV, IVC, and BEOL. The detail of the TJ IVC is shown in the inset.

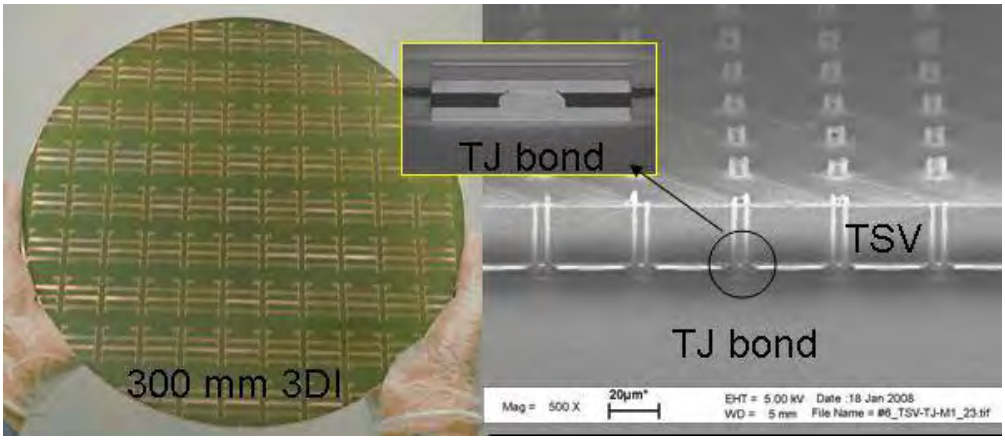


Figure 7: a) A 300 mm TJ bonded 3DI wafer, b) The TSV-IVC-BEOL via chain, c) TJ Structure Details, all from (17).

Summary

The transfer and join (TJ) technology advances are reviewed from MCM-D to recent 300 mm wafer level 3DI. The versatility of the technology can enable high density fine pitch 2D and 3D integrations at wafer-wafer (W2W) and chip-wafer (C2W) levels for "More Moore" and "More than Moore" scaling in the areas of device, MEMS, opto-electronics, hybrid devices. These advances can ultimately bring SOC and SIP together, achieving SOC performance with SIP versatility for the future CMOS.

Acknowledgments

The author wishes to thank Drs. S. Purushothaman, S. Koester and W. Haensch for their critical review and comments of this manuscript.

References

(1) G. Moore, Electronics, V38, pp114, 1965
(2) 2007 ITRS, http://www.ITRS.org/
(3) T.C. Chen, ICSICT2006, pp4-7.
(4) G.G. Shahidi, IEEE-CICC 2007, pp413-416.
(5) G. Amdahl, AFIPS Conference Proceedings, (30), pp.483-485, 1967.
(6) J.E. Fritts and E.D. Chamberlain, proc. 35th Simulation Symposium, 2002. Page(s):352 - 362
(7) P. Jacob et al, IEEE Design&test, V22,No.6, 2005, pp540-547
(8) S.J. Koester et al, IBM J.of R&D V52, N6, Nov. 2008, pp583-595
(9) P.S. Andry, et al, Proc. of the 56th ECTC, pp. 831-837, 2006.
(10) P. R. Morrow et al, *IEEE Elect. Dev. Lett*. 27, 335-337 (2006).
(11) H. B. Pogge et al., Proc. AMC'2001, pp.129-136,2001;
(12) R. Yu, proc. ICEPT2008, Shanghai, July 28-31, pp.1-10.
(13) C. Narayan, et al., IEEE trans. Comp. Pkg. Mfg. Tech., Part B, Vol 18, Issue 1, 1995, p42
(14) K. W. Guarini, et al, *IEDM Tech. Digest*, 943-946 (2003).
(15) R. Yu, proc. VMIC2007, p223-230.
(16) M. Despont, et al; TRANSDUCERS, Solid-State Sensors, Actuators and Microsystems, 12th International Conference on, 2003,Volume 2, 8-12 June 2003 Page(s):1907 - 1910 vol.2
(17) F. Liu, et al, IEDM 2008.

Ex. 1016, Page 884

ECS Transactions, 18 (1) 857-862 (2009)
10.1149/1.3096546 © The Electrochemical Society

Exploratory NEMS-CMOS Hybrid Devices for Post CMOS era

B.H. Lee

Department of Material Science and Engineering,
Gwangju Institute of Science and Technology
Oryong 1, Buk-gu, Gwangju, Korea, 500-712, bhl@gist.ac.kr

Si device technology is facing the limit of physical scaling for several reasons such as a power management, an increasing cost of physical scaling, and a variability of nano scale devices. Furthermore, the physical scaling alone does not provide strong benefits on performance as in the past technology generations due to increasing parasitic components (capacitance and resistance) and rising chip temperature. Yet, the demands for increased functionality still exist. Thus, to continue increasing the level of integration, it is important to find a way 1) to control the power consumption and dissipation, 2) to increase the utilization of existing devices within the chip, and 3) to improve the efficiency of interconnect. Hybridization of nano electro mechanical switch (NEMS) with CMOS technology has been proposed to reduce the power consumption in some circuit modules by one to two orders of magnitude with a minimal cost at performance. NEMS-CMOS hybrid device technology appears to be an attractive option in terms of power management, but there are not enough learning on true nano scale NEMS device, which can be readily integrated with Si CMOS technology. In this talk, various options for NEMS-CMOS hybridization and associated challenges in materials and devices will be reviewed.

1. Introduction

Successful execution of physical scaling which enabled Moore law during last four decades is facing serious challenges. Physical scaling of critical dimension is getting more expensive as it requires EUV lithography and complex integration schemes incorporating various materials and strain engineering. Also, one of the key factors that have enabled the scaling of MOS devices, the scaling of gate dielectric, has been off track since 2002.[1] Thus, traditional meaning of Moore law has been redefined as a scaling of performance instead of the scaling of chip size oe critical dimension.

In addition to the challenges in the physical scaling, CMOS technology has many other road blocks in future, especially when it targets beyond Tera-bit scale integration. Power consumption in leading edge CPU chip will continuously increase due to the increasing number of devices per chip and the increasing demands for functionality[2]. Increasing power consumption at interconnect which will exceed the power consumption from frontend devices in near future is another serious problem. Variability at nano scale devices is also another factor that degrades the power consumption and limit supply voltage scaling (V_{DD}) to solve the power consumption problem.

Ex. 1016, Page 885

There are many technologies being investigated to mitigate the problems due to increasing power consumption and variability at nano scale deices. Various circuit/architectural technologies such as local hibernation, multicore scheme etc have been actively pursued.[3] At a device level, devices with faster on-off characteristics such as tunneling FET or impact ionization FET are proposed as a way to reduce $V_{DD.}$[4] Yet, impact ionization devices (I-MOS) are suffering from reliability and variability problems due to a high energy carrier injection from source side.[4,5] Tunneling devices provides a more stable operation window, but a current drivability is limited and the gain from slight improvement in swing is somewhat limited as a fundamental solution for future scaling.[6] For example, an ideal subthreshhold swing of CMOS, 60mV/dec, can be improved to 20-40mV/dec using various tunneling FET devices. However, these improvements can only be translated into 0.1V-0.2V V_{DD} reduction assuming 5 decades of I_{on}/I_{off} ratio. Thus, tunneling FET can only provide technologies for 1-2 generation.

Considering that there are no alternative non-Si based device technologies that can match the silicon based CMOS technology in terms of mass integration and production, novel technologies that can extend the life of silicon based device technology for next one or two decades are urgently needed. Key features required in those technologies are following. 1) Reduce the power consumption by at least 1-2 orders of magnitude, 2) Increase the utilization of existing devices within the chip even without scaing, and 3) Improve the efficiency of interconnect by altering signal path (like 3D interconnect) or sharing of interconnect with other circuit module.

Above challenges require various novel approaches at all levels of semiconductor technology including process, device (Frontend and backend), circuit and architecture. In this talk, an approach to reduce the power consumption of CMOS device by combining nano electro mechanicel switch (NEMS) with CMOS devices will be introduced.

2. Nano Electro Mechnical Switch (NEMS)

Mechanical switch has been studied for a long time, primarily for the applications in sensors, RF devices and memory applications. Mechanical motion of electrode (or channel) alters the conduction path through the electrode or channel region. Gate electrode may be clamped in one side (cantilever type) or two sides (type).

Fig.1 shows a collection of various mechanical switches demonstrated since 1980. From macro scale devices shown in Fig.1 (a) to true nano scale tweezer type carbon nanotube switches, the dimension has been continuously reduced. Challenges in the actual implementation of NEMS technology are following. 1) Too high operation voltage to pull in the electrode, 2) limited switching speed (>> 1nsec), 3) Control of nano scale gap, 4) Reliability of NEMS (mechanical failure, stiction). In addition, to be compatible with CMOS technology, actual dimension of NEMS needs to be scaled down to the size of base devices (e.g. Si-SRAM).

The switching speed of NEMS is determined by the geometry of electrode, Young's modulus of electrode materials and its clamping structures, the range of electrode motion (gap between electrodes or electrode and channel). Operation speed of NEMS is approximately given by the resonant frequency and elastic constant of electrode material.[7].

Ex. 1016, Page 886

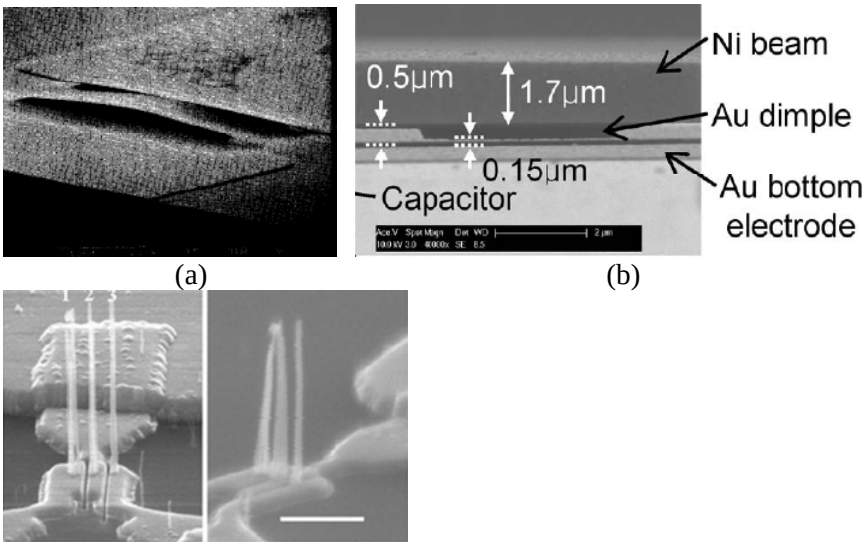


Figure 1. (a) micro scale MEMS with double clamp[8], (b) NEMS with TiN electrode[9], (c) NEMS with CNT electrodes.[10]

$$t_o \approx \frac{0.15 \cdot l^2}{d \cdot \sqrt{E/\rho}} \qquad [1]$$

Here l, d, E, ρ are the gate length, electrode thickness, Young's modulus, density respectively. Note that this equation assumes the displacement of electrode is much smaller than actual oscillation amplitude.

Fig.2 (a) shows the switching speed of various electrode materials. Due to its extremely thin thickness, the resonant frequency of two layer graphene (0.68nm) is relatively lower than other materials. However, gate bending at the given gate length shows that silicon and TiN are too stiff to be practical at this dimension (< 100nm). Max. gate bending was approximately calculated using max strain value; $\Delta d = 2\epsilon_s l/3$.

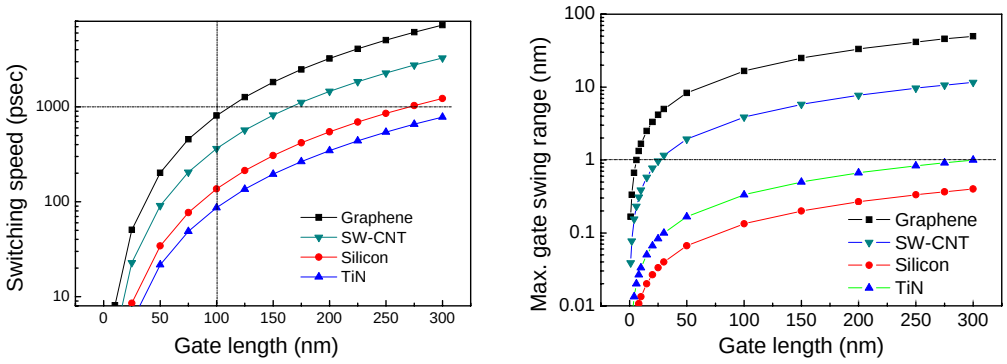


Figure 2. (a) Switching speed of cantilever consists of various materials, (b) maximum gate swing (=deflection) as the given gate length.

TABLE I. Parameters used to calculate curves shown in Fig.2 (a) and (b)

Parameters	Graphene	CNT	Silicon	TiN
Density ($g.cm^3$)	2.2	2.6	2.3	3
Young's Modulus(GPa)	1000	1200	170	550
Max. strain ϵ_S (%)	25	5.8	0.2	0.5
Thickness (nm)	0.68	1.5	20	20

Considering that the implementation of NEMS into CMOS may take 2-3 generations further down in ITRS roadmap (corresponding to 6-10 years), the dimension of NEMS device should be much smaller than 100nm. Unfortunately, most of material studied for NEMS applications does not work at this physical dimension. As shown above, a realistic gap that can be reliably controlled with current technology might be not much smaller than 10nm. In this dimension, materials like TiN or Si cannot be bent down by 10nm with 100nm length cantilever as shown in Fig.2 (a) and (b). Thus, unless the implementation of NEMS can be done in vertical direction, CNT or graphene are needed as electrodes materials for NEMS-CMOS hybrid devices to realize a true nano scale NEMS devices. Unfortunately, the fabrication process for both CNT and graphene are not mature enough to be applied to a large scale wafer integration and this is the biggest challenge in this area.

3. NEMS-CMOS hybrid devices

Even though the actual implementation of NEMS device might be delayed due to the availability of base materials and its integration processes, various concepts to implement NEMS into Si-CMOS technology are actively studied at circuit or architectural level.

Fig.3 shows various schematics of NEMS circuit modules discussed in the literature. An inverter consists of two NEMS(Fig.3 (a) and (b)) [11,12,13] or a SRAM consist of 4 NEMS and 2 Si MOSFET[14] have been proposed. These circuits modules can save the power consumption by orders of magnitude using very steep turn on characteristics and low V_{th} even though these benefits are only demonstrated at theoretical model up to now. NEMS-CMOS hybrid technology can be implemented within FEOL device or can be combined with NEMS embedded in BEOL interconnect.

While all of these ideas are using one dimensional and single level switching action (Up and down motion or lateral motion of cantilever), it might be possible to use NEMS for multi level switching, parallel processing of multiple signal and also as a switch to disconnect power and ground line during hibernation to minimize the charge dissipation within the interconnec.[15] More design and devices concepts need to be further developed to mitigate the weakness of NEMS devices such as a slow switching speed and a reliability problem and to maximize the benefits of an ultimate leakage current insulation and a high speed turn-on characteristics. In addition, there are several ideas to introduce the concept of mechanical switching devices at various stage of Si device integration. NEMS can be introduced at M1 level or over metal interconnect using 3D integration technology. These BEOL NEMS devices can be made using more conventional materials such as Al, TiN, TaN or Cu.

Ex. 1016, Page 888

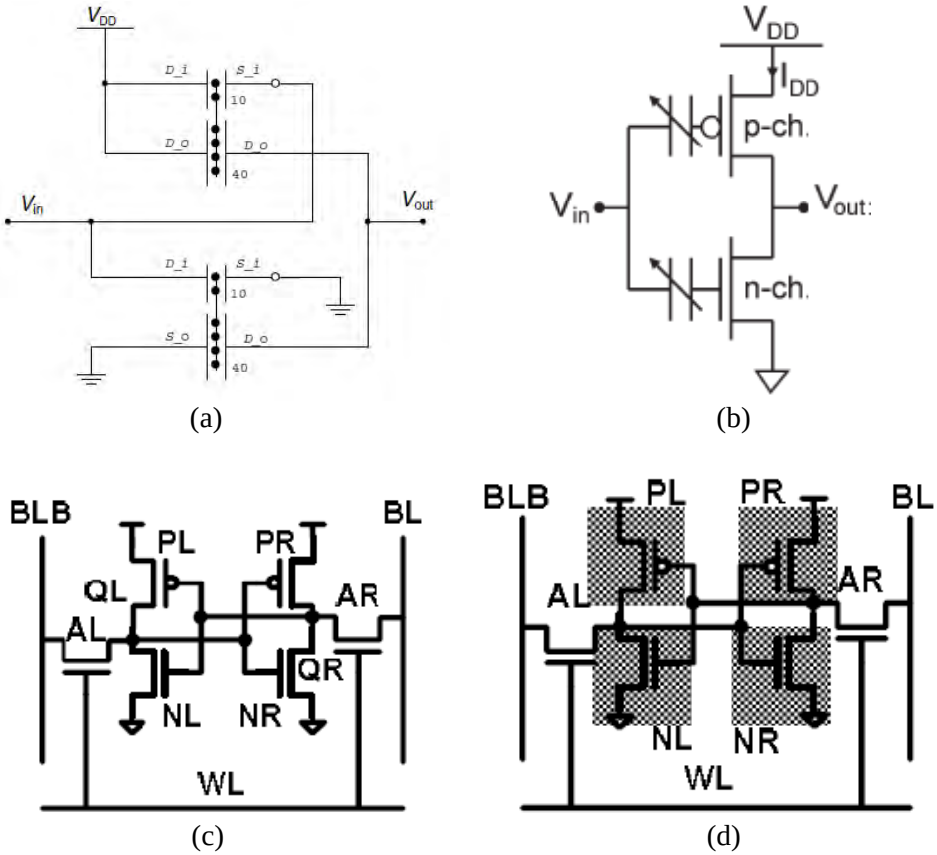


Fig.3 a) Inverter consists of two coupled nano-electromechanical single electron transistors(NEMSET)[11], b) Schematic of the complementary SGFET inverter[12], c) conventional SRAM, d) NEMS-CMOS hybrid SRAM[14]

Also, NEMS device can be used as memory devices using its hysteretic on-off curve.[10,16] Theoretically, NEMS memory device can achieve reasonable operation speed (read/write < 1nsec) which can be used as M1 level memory integrated with logic devices or as a universal memory. However, to achieve Tera-bit scale integration for more general applications, these memory devices need to use graphene or CNT electrode or go to alternative path, multiple stacking of NEMS devices in vertical direction.

4. Summary

NEMS-CMOS hybrid devices and circuit technology has a strong potential to reduce the power consumption of current CMOS technology by orders of magnitude and can become a key enabler to continue the use of Si based manufacturing technology for upcoming decades. However, the key element to realize the technology, highly robust NEMS device itself, has not been successfully demonstrated yet. Development of electrode materials with a high mechanical strength, a minimal stiction problem and good reliability characteristics are the most important homeworks for NEMS-CMOS hybrid technology.

Ex. 1016, Page 889

Acknowledgments

This works was partially supported by DASAN project through a grant provided by Gwangju Institute of Science and Technology in 2008.

References

1. B.H.Lee *et al.*, *Materials Today*, **9**, 32 (2006).
2. B.H.Lee *et al.*, *IEEE Trans. on Electron Device*, **55**, 8 (2008).
3. B. Nicholic, *IEEE Trans. Electron Dev.*, **55** (1), 12 (2008).
4. C. Charbuillet *et al.*, in *Tech. Dig. of IEDM* (2006).
5. F. Mayer *et al.*, *IEEE Tans. on Electron Dev.*, **55** (6), 6 (2008).
6. K. Boucart and A. M. Ionescu, *IEEE Tans. on Electron Dev.*, **54** (7), 9 (2007).
7. S. Fujita *et al.*, *IEEE Trans. on Circuits and Systems*, **54** (11), 8 (2007).
8. B. Halg, *IEEE Tran. on Electron Device*, **37** (10), 7 (1990).
9. W. W. Jang *et al.*, *IEEE Tran. on Electron Device*, **55** (10), 5 (2008).
10. J. E. Jang *et al.*, *Applied Physics Letters*, **87**, 163114 (2005).
11. R. H. Blick *et al.*, *New Journal of Physics*, **9** (2007).
12. K. Akarvardar *et al.*, *IEEE Trans. on Electron Devices*, **55** (1), 48-59 (2008).
13. D. Tsamados *et al.*, *Solid-State Electronics*, **52** (9), 1374-1381 (2008).
14. H. F. Dadgour and K. Banerjee, in *DAC* (San Diego, CA, 2007), p. 306.
15. A. M. Ionescu, International Symposium on VLSI Technology, Systems, and Applications, Proceedings, 72-73, (2008).
16. D. Molinero *et al.*, Proceedings of the IEEE International Conference on Micro Electro Mechanical Systems (MEMS), 685-688, (2008); W. W. Jang *et al.*, *Solid-State Electronics*, **52** (10), 1578-1583 (2008); M.-S. Kim *et al.*, in *ISDRS* (College Park, MD, 2007).

ECS Transactions, 18 (1) 863-868 (2009)
10.1149/1.3096547 © The Electrochemical Society

Evaluation of InAs QWFET for Low Power Logic Applications

Edward Yi Chang[a,] ,Chien-I Kuo [a] ,Heng-Tung Hsu[b] ,Yasuyuki Miyamoto[c] ,Chia-Ta Chang[a] and Chien-Ying Wu[a]

[a]Department of Materials Science and Engineering, National Chiao-Tung University, Hsinchu, Taiwan 30010, R.O.C.
[b] Department of Communications Engineering, Yuan Ze University, Chungli, Taiwan 32003, R.O.C.
[c] Department of Physical Electrons, Tokyo Institute of Technology, Tokyo 152-8552, Japan.

An eighty nanometer gate length InAs quantum well field effect transistors has been fabricated and the digital characteristics were evaluated. The devices show a drain-source current of 1015 mA/mm and a peak g_m of 1920 mS/mm at V_{DS} = 0.5 V. The f_T and f_{max} are 340 GHz and 220 GHz at V_{DS} = 0.4 V, respectively. A low delay time ($C_{total}V/I_{ON}$) of 0.54 ps was also achieved. These excellent results indicate the InAs devices are the great potential option for the future low-power logic applications at post-Si CMOS technology.

Introduction

For future high-speed and low-power digital applications, Quantum Well Field Effect Transistors (QWFETs) with InGaAs channel materials of high indium concentration is an attractive transistor technology at post-Si-CMOS era due to the high electron mobility and high saturation velocity of the InGaAs material (1-4). In fact, the excellent RF-performance with world-record current-gain cutoff frequency of 628 GHz has been demonstrated by InAlAs/InGaAs QWFETs on InP substrate (5).

According to the latest ITRS summer conference, III-V compound semiconductors are considered as emerging alternative channel materials to extend moore's law (6). In generally, the InAs has high electron mobility of 40,000 cm^2/Vs at room temperature, high electron peak velocity, low electron effective mass ($0.023/m_0$), reasonable energy bandgap (0.36 eV) and high quantum confinement in the channel region (7). Therefore, it is considered suitable channel material for high speed electronic application in 22 nm generation and beyond.

In this study, we report the fabrication and evaluation of 80 nm InAs-channel QWFETs for high speed and low power applications beyond the Si CMOS.

Experimental

The QWFET epitaxial structures were grown by molecular beam epitaxy on a 2-inch semi-insulating InP substrate. Fig. 1 (a) and 1 (b) represent the detailed epitaxial layer of InAs channel device and the simulated energy band diagram. The calculated sheet charge density is 2.1×10^{12} $/cm^2$ and the large conduction band offset for InAs channel is suitable for QWFETs applications.

Mesa etching was performed to provide isolated active areas by using phosphoric/H_2O_2/H_2O. Low ohmic metallization was formed and showed a specific contact resistance of ~ 1.5×10^{-7} Ω cm^2 after rapid thermal annealing at 250 ^oC for 25 sec in a forming gas condition. A tri-layer resist system of ZEP-520/PMGI/ZEP520 was used for the E-Beam lithography and the gate length of the T-shaped gate was 80 nm. Gate recessing was carried out in the mixture of succinic acid/H_2O_2 solution. Besides, the recess must be controlled precisely which is the most critical step in fabricating the QWFETs (8). After gate recess, the Pt (12nm)/Ti (60nm)/Pt (80nm)/Au (180nm) were deposited as Schottky gate metal and lifted off using ZDMAC remover (ZEON Corporation) to form 80-nm T-shaped gate. The scanning electron microscope (SEM) images of the device and the T-shaped gate are shown in the Fig. 2 (a) and (b), respectively.

Result and Discussion

The drain–source current I_{DS} versus drain–source voltage V_{DS} curves and the transconductance g_m versus gate–source voltage V_{GS} curves of the 80-nm gate InAs QWFETs are shown in Fig. 3 (a) and (b) respectively. The 2 × 50-μm-wide device exhibited good pinchoff characteristics, and the device is a depletion mode with a threshold voltage (V_T) of −0.81 V. The current density I_{DS} was 1015 mA/mm. The maximum g_m of the device at a V_{DS} of 0.5 V was 1920 mS/mm. The high transconductance was due to low contact resistances, high carrier concentration and superior electron transport properties in the InAs channel. The result indicates that the InAs-channel QWFETs can be operated at low bias condition to reduce dc power dissipation while maintaining high performance. Additionally, these devices performing faster switching speed and better on-current performances than Si transistors will play an important role on future logic devices.

The S-parameters for the InAs QWFETs were measured from 2 to 80 GHz using on-wafer 100-μm-pitch ground–signal–ground probes. The current gain cutoff frequency was defined by extrapolating the H_{21} curve at a slope of 20 dB/dec. Fig. 4 (a) shows the frequency at different V_{DS} as a function of current density and fig. 4 (b) shows the f_T dependence of the current gain H_{21}, the power gain "MAG/MSG," and the unilateral gain U of the InAs devices measured at V_{DS} = 0.4 V and V_{GS} = − 0.45 V. The current gain cut-off frequency f_T and the maximum oscillation frequency f_{max} obtained for the 2 × 20 μm QWFETs were 220 and 340 GHz respectively.

An important figure of merit for high speed operation of logic transistors, the intrinsic gate delay ($C_{total}V/I_{ON}$), has also been investigated. According to the definition in (9,10), C_{total} is the total gate capacitance extracted from the scattering parameter. V and I_{ON} are the applied drain voltage and on-state current respectively. For InAs devices, the calculated intrinsic gate delay of 0.54 ps was achieved. The comparisons of 80 nm InAs QWFETs, 200 nm InSb QWFETs and advanced 40 nm Si MOSFET on gate delay and I_{ON}/I_{OFF} ratio were plotted in the Fig. 5 (a). The InAs QWFETs exhibits a far better I_{ON}/I_{OFF} ratio than the InSb QWFETs meanwhile attaning the subpicosecond intrinsic speed. Furthermore, the gate delay of InAs or InSb QWFETs are much dependent on the I_{ON}/I_{OFF} as compared to the Si MOSFETs. This is mainly due to the off-state current of QWFETs which is determined by the gate leakage current. It was limited resulting from band-to-band tunneling (BTBT) in such small energy band gap and low effective conductivity mass (m*). Therefore, the use of a high-k gate dielectric between the metal

Ex. 1016, Page 892

gate and the III-V device layers are needed to eliminate the gate leakage and potentially improve the I_{ON}/I_{OFF} ratio (11).

Fig. 5 (b) shows the cutoff frequency versus dc power consumption of InAs QWFETs in this study compared to reported 80 nm Si NMOSFETs. We can conclude from the plot that InAs QWFETs show > 3 × gain in speed performance for the same power compared to Si nMOSFETs, indicating the great potential of InAs QWFETs in low-power and high-speed logic applications. It can be also seen from Fig. 5 (c) that InAs QWFETs exhibit a lower gate delay than the advanced Si nMOSFETs comparing at the same gate length.

Conclusion

The InAs QWFETs had been fabricated successfully and benchmarked against Si nMOSFETs for high-speed and low-power digital applications. The high drain current density of 1015 mA/mm and transconductance of 1920 mS/mm at low drain voltage V_{DS} = 0.5 V were obtained due to the high electron mobility in the InAs channel. In addition, comparison of advanced Si nMOSFETs, the InAs QWFETs exhibited superior performance on gate delay time and other characteristics. With further improvement of I_{ON}/I_{OFF} ratio and process, InAs-based devices could be the candidates for future nanoelectron application beyond the Si-CMOS.

Acknowledgments

This work was supported by the National Science Council and the Ministry of Economic Affairs, Taiwan, R.O.C., under Contract NSC 97-2221-E-009-156-MY2 and Contract 96-EC-17-A-05-S1-020. Part of this work was supported by the ‘‘Nanotechnology Network Project’’ of the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

References

1. D. H. Kim, J. A. Alamo, J. H. Lee, and K. S. Seo, *IEEE Transactions on Electron Devices*, **54**, 2606 (2007).
2. C. Y. Chang, H. T. Hsu, E. D. Chang, C. I. Kuo, Y. Miyamoto, S. Datta, M. Radosavljevic, and G. W. Huang, *IEEE Electron Device Letters*, **28**, 856 (2007).
3. R. Chau, B. Doyle, S. Datta, J. Kavalieros and K. Zhang, Nature Materials, **6**, 810 (2007)
4. C. I. Kuo, H. T. Hsu, E. D. Chang, C. Y. Chang, Y. Miyamoto, S. Datta, M. Radosavljevic, G. W. Huang, and C. T. Lee, *IEEE Electron Device Letters*, **29**, 290 (2008).
5. D. H. Kim and J. A. del Alamo, *IEEE Electron Device Letters*, **29**, 830 (2008).
6. International Technology Roadmap for Semiconductors, in ITRS Summer Conference 2008, San Francisco, CA.
7. S. Takagi, T. Irisawa, T. Tezuka, T. Numata, S. Nakaharai, N. Hirashita, Y. Moriyama, K. Usuda, E. Toyoda, S. Dissanayake, M. Shichijo, R. Nakane, S. Sugahara, M. Takenaka, and N. Sugiyama, *IEEE Transactions on Electron Devices*, **55**, 252 (2008).
8. J. H. Oh, Y. H. Baek, B. O. Lim, S. W. Moon, S. J. Lee, J. K. Rhee, I. S. Hwang, and S. D. Kim, *Journal of The Electrochemical Society*, **154**, 541 (2007).

Ex. 1016, Page 893

9. R. Chau, S. Datta, M. Doczy, B. Doyle, B. Jin, J. Kavalierous, A. Majumdar, M. Metz, and M. Radosavljevic, *IEEE Transcations On Nanotechnology*, **4**, 153 (2005).
10. J. Guo, A. Javey, H. Dai, and M. Lundstrom, *IEDM Tech. Dig.*, 2004, pp.703-706.
11. R. Chau, J. Brask, S. Datta, G Dewey, M. Doczy, B. Doyle, J. Kavalieros, B. Jin, M. Metz, A. Majumdar, and M. Radosavljevic, *Microelectronic Engineering*, 80, 1 (2005)

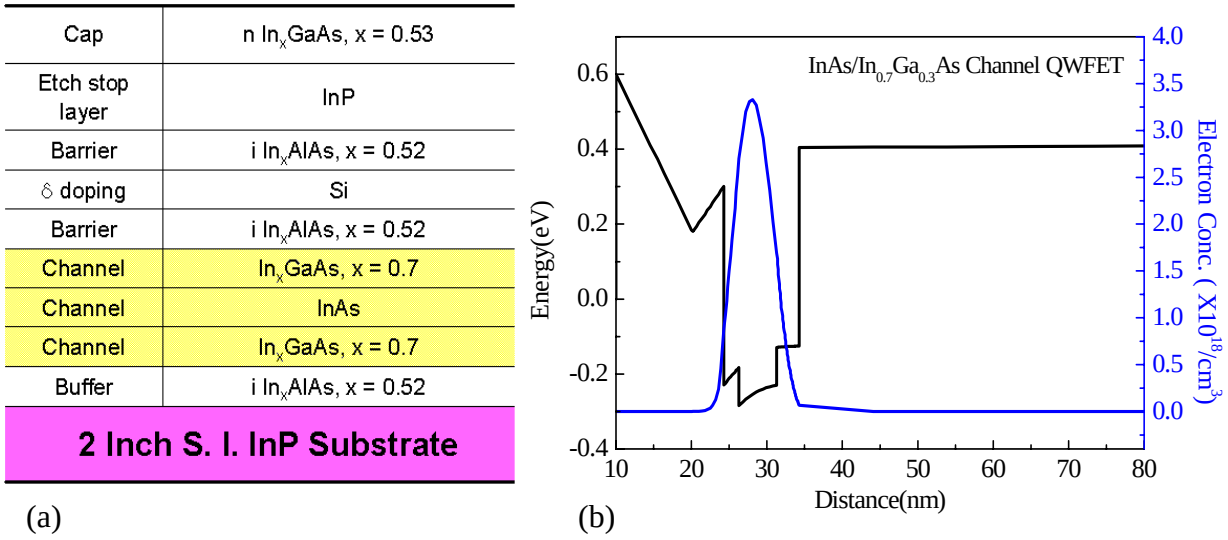


Fig. 1 (a) InAs QWFETs Epitaxy layer structure in this study and 1 (b) Energy band diagram simulation of InAs QWFETs

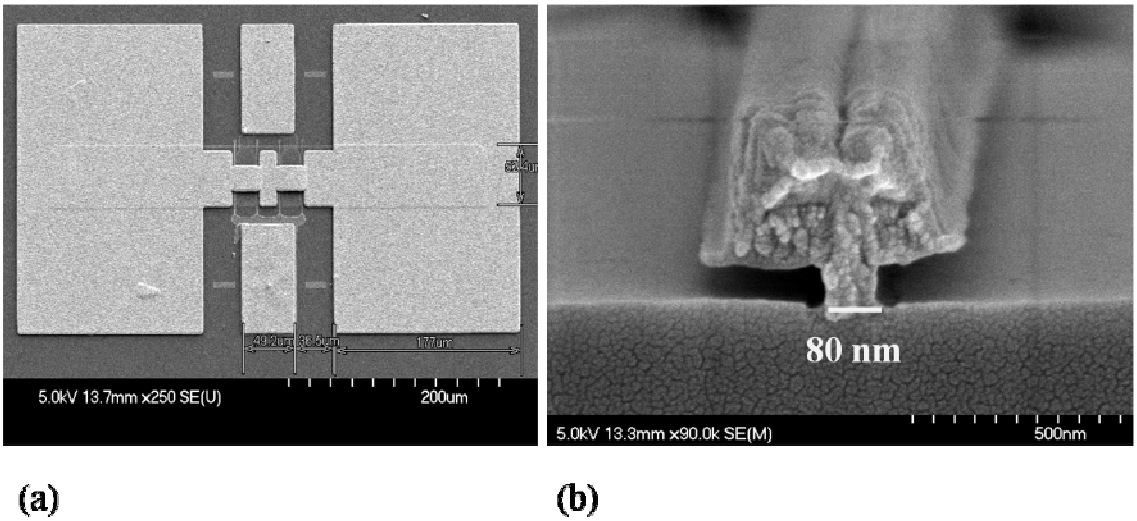


Fig. 2 (a) SEM image of the device with total gate peripheries of 100 μm and 80 nm gate length and 2 (b) the cross-sectional SEM image of the 80-nm T-shaped gate InAs-channel QWFET of interest before SiNx passivation.

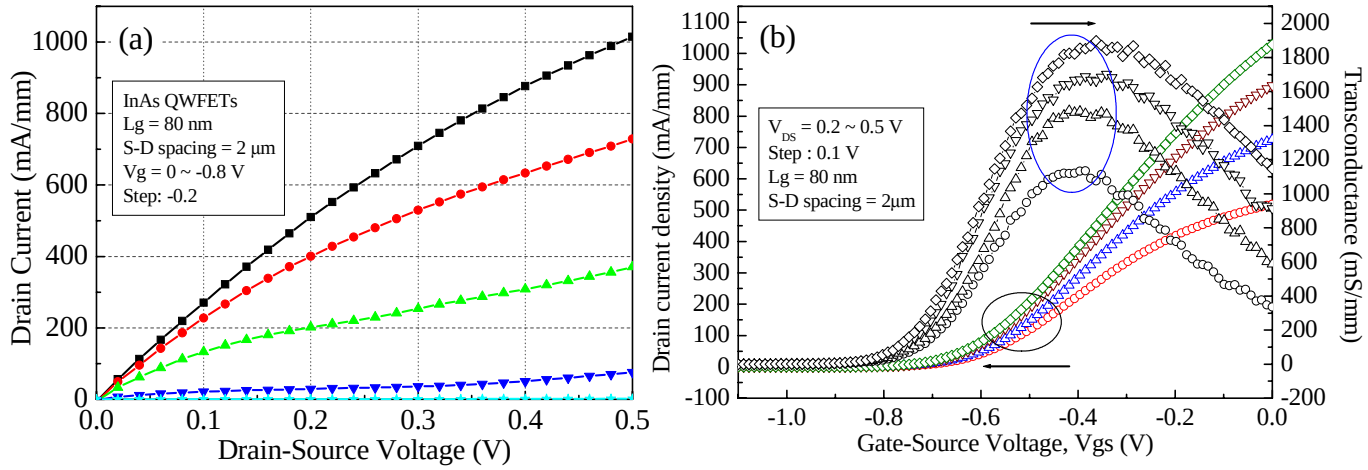


Fig. 3. (a) I_{DS} versus V_{DS} curve of 80 nm InAs QWFETs (b) Current-voltage characteristics at different V_{DS} of 80 nm InAs QWFETs.

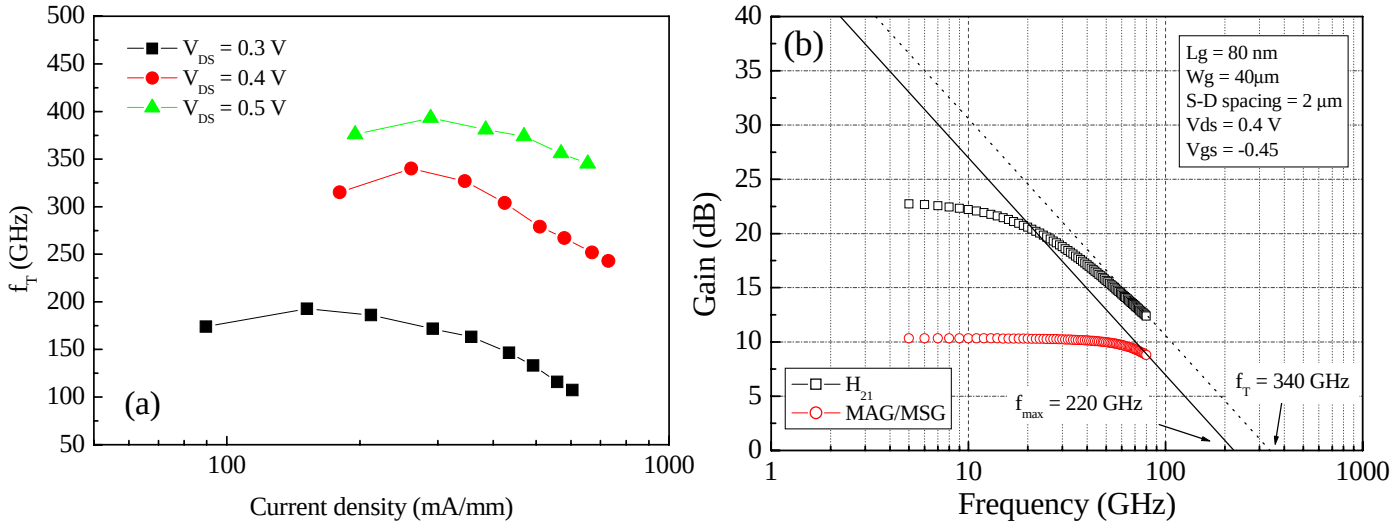


Fig. 4 (a) f_T of 80 nm InAs QWFETs as a function of current density at different V_{DS} (b) Frequency dependence of the current gain H_{21}, and the power gain MAG/MSG of the 80 nm InAs QWFETs. The frequency range was from 5 to 80 GHz, and the device was biased at V_{DS} = 0.4 V and V_{GS} = - 0.45 V.

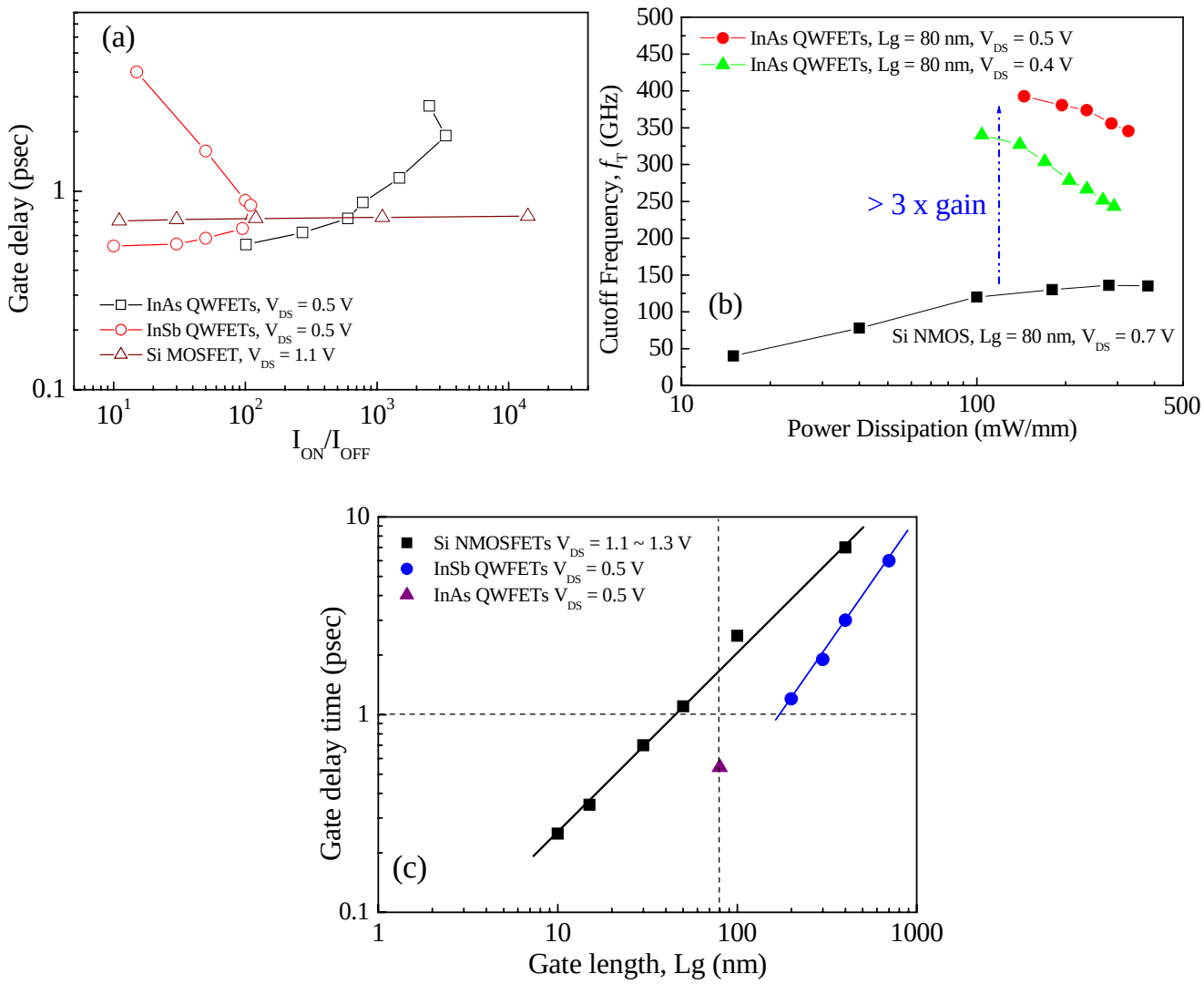


Fig. 5 (a) Comparison of gate delay and I_{ON}/I_{OFF} among 80 nm InAs QWFETs, 200 nm InSb QWFETs and 40 nm Si MOSFETs. (b) Cutoff frequency of InAs QWFETs and 80 nm Si MOSFETs as a function of the power dissipation. (c) Gate delay of InAs, InSb QWFETs and Si NMOSFETs as a function of gate length.

ECS Transactions, 18 (1) 869-874 (2009)
10.1149/1.3096548 © The Electrochemical Society

Translucent Polymer Solar Cells

G. M. Ng, C. Vijila, W. P. Goh, F. R. Zhu*

Institute of Materials Research and Engineering,
A-Star (Agency for Science, Technology and Research),
3 Research Link, Singapore 117602, Singapore
*Email: fr-zhu@imre.a-star.edu.sg

The recent developments on the construction of translucent cathodes for the application of tandem or multi-junction organic solar cells are reviewed in this report. The translucent cathode plays a major role in determining the efficiency of translucent solar cells by controlling the intensity of incident light absorption at the photoactive layer of each cell. As a result, there are varieties of translucent cathodes based on conductive polymer, thin metallic layer, and transparent conducting oxides (TCO) have been developed by various research groups. A translucent cathode based on a combination of thin metallic (i.e. calcium and silver) and transparent conductive oxide layers (indium tin oxide (ITO)) is developed in our group and the advantages of this translucent cathode for organic solar cell application are presented. In addition, the experimental results obtained from polymeric solar cells, based on poly(3-hexylthiophene) (P3HT): 1-(3-methoxycarbonyl)-propyl-1-phenyl-(6,6)C61 (PCBM) blends, and ITO-based translucent cathode, are also presented.

I. Introduction

In recent years, the developments of polymeric photovoltaic materials and devices are attracting significant research interest due to their potential for light weight, flexible, inexpensive solar cell applications. The development of efficient and stable photovoltaic cell is the major challenge with the current organic solar cell (OSC) technology. Significant improvements on the power conversion efficiency of polymeric solar cells have been achieved using the bulk heterojunction (BHJ) between an electron donating conjugated polymer and an electron accepting fullerene derivative (1-2). However, the device performances with single photo-active layer solar cells have not reached the efficiency level required for practical applications. As a result, significant research effort is focused on the development of tandem or multilayered organic photovoltaic cells with an aim to utilize the solar energy effectively. In tandem or multilayered solar cells, two or more subcells, with non-overlapping spectral response, are stacked in order to cover the broader solar spectrum compared to the single junction cells. These subcells are separated with suitable translucent electrode which allows light transmission into the subsequent photoactive layers (3-5). Therefore, it is very important to investigate the light transmission properties of the translucent electrodes by identifying a suitable electrode material and design in order to obtain the photocurrent match in the stacked solar cells. In addition, the solar cells with translucent cathode have a see-through feature which makes

Ex. 1016, Page 897

them suitable for application in architectural or power-generating windows and greenhouses.

It has been shown in the literature that the translucent solar cells can be fabricated using variety of translucent cathodes such as conducting polymer (6), thin metallic layers (7-8), transparent conducting oxide (TCO) layer (9-11) and a combination of thin metallic and transparent conducting oxide layers (12). It has been observed that the polymeric solar cells based on APFO-3: PCBM blend as an active layer and VPP: PEDOT layer as a translucent cathode shows the power conversion efficiency less than 0.01% (6). The very low power conversion efficiency of the polymeric solar cell with conducting polymer as a translucent cathode was ascribed to the poor bulk conductivity of the polymer electrode.

The translucent organic solar cell with a combination of few thin metal layers as a cathode has been introduced by Shrotriya et al (7, 8) in order to improve the conductivity and transparency. It was observed that the translucent cathode based on lithium fluoride (LiF)/aluminium (Al)/gold (Au) showed the best transparency due to the wettability enhancement of the deposition of Au onto LiF/Al surface. The translucent polymeric solar cell showed a PCE value of about 1.1% with the device structure ITO/PEDOT:PSS/MEH-PPV:PCBM(70nm)/LiF(1nm)/Al(2.5nm)/Au(12.5nm). However, the transparency of the complete device is not investigated in detail. Hadipour et al. have also reported a similar system of translucent OSC using Sm (3nm)/Au (12nm) as the translucent cathode and this device exhibits the PCE of about 3% using a thick P3HT:PCBM (250nm) photoactive layer (9). However, the transparency of this device is expected to be much low due to the thick photoactive layer and the large reflection and absorption by the 15nm metal cathode used. This type of translucent OSC has restriction in increasing the thickness of the thin metal layer used in the cathode because the transmittance will be decreased drastically.

The translucent electrode using a TCO as the cathode, have been developed in order to overcome the strong absorption of thin metal layer. ITO is commonly recognized as a high transparent electrode in the visible range with desired conductivity. Zaus et al. reported a translucent organic photodiodes using thin Ca/Ag and LiF/ITO as the translucent cathode (10). It has been observed that the Ca/Ag cathode was affected by large interface reflection and strong absorption in the visible range that leads to low device transparency.

Kawano et al. reported a translucent OSC using ITO as the transparent cathode without using any buffer layer (11, 12). The device with the structure of glass/ITO/PEDOT:PSS/MDMO-PPV:PCBM(80nm)/ITO(20nm) showed a reduction in the open circuit voltage (Voc) as compared to glass/ITO/PEDOT:PSS/MDMO-PPV:PCBM(80nm)/Al(100nm) due to the symmetrical electrodes used in the translucent OSC since Voc is strongly influenced by the work function difference of anode and cathode used in the device, demonstrating that ITO is more stable and rigid as an interlayer in tandem structure as compared to the very thin Ag layer.

II. Polymeric solar cells with a combination of thin metal and transparent conductive oxide layer as a translucent cathode

It is generally accepted that the implementation of TCO as a cathode in translucent OSC increases the light transmittance T(λ) of the device, but most of the TCO is deposited by sputtering process which generates damages to the underlying material especially organic materials. Therefore, it is recommended to introduce a buffer layer at the polymer/cathode interface, which can be a thin metal or polymer layer. The buffer layer has to be electrically favored to the device operation, and also be able to withstand any damages caused during TCO sputtering process. Morimune et al. reported a translucent organic photodetectors using an amorphous carbon nitride (a-C:N) as a buffer layer between the organic and ITO laser, that prevents the damages created on the organic layer during ITO sputtering process (13). However, the focus of the manuscript is mainly on the sensing characteristics of the photodetector and the I-V characteristics of these detectors indicates that the cathode fabrication process needs to be optimized further for the translucent organic solar cell application. Bailey-Salzman et al. demonstrated a translucent OSC using bilayer compound cathode of Ag/ITO (14). It has been observed that the light transmission properties of the entire device can be varied by adjusting the thickness of the Ag layer. The average light transmission of 26% with PCE of 0.62% was obtained for 10nm thick Ag layer whereas the light transmission of 60% with PCE 0.28% was achieved without the Ag layer. The low power conversion efficiency without the Ag layer was attributed to the relatively low light absorption in the active layer of the translucent OSCs. However, the variation of the ITO thickness to the performance of the translucent OSCs was not discussed.

We have applied optical admittance analysis to calculate the integrated absorptance, A(λ) and the T(λ) of the translucent OSC with the structure of ITO/PEDOT:PSS/P3HT:PCBM(75nm)/Ca(10nm)/Ag(10nm)/ITO prior to the device fabrication. It is observed that the translucent OSC with an optimized Ca(10nm)/Ag(10nm)/ITO cathode is superior to a thin transparent metal layer as cathode. A translucent OSC with a thin multilayered metal cathode, e.g., Ca(10nm)/Ag(10-15nm), causes a large internal reflection at the metal/air interface due to its high refractive index. The use of Ca(10nm)/Ag(10nm)/ITO cathode enhances the electrical and optical properties of ITO layer that enable it serves as a transparent conducting index matching layer to enhance the total light transmission. In addition, the upper ITO contact also acts as a passivation layer to protect the underlying functional organic layers in OSCs from any possible moisture attack from the environment.

The effect of the ITO cathode thickness on the T(λ) and A(λ) of P3HT:PCBM (75nm) in translucent OSC were also analyzed as depicted in Figure 1. The calculated results indicated that a maximum T(λ) through translucent OSC can be achieved at an ITO layer thickness of 50nm as shown in Figure 1 (solid curve), whereas the maximum light absorption occurs at an ITO cathode thickness of 80nm (dashed curve). The required ITO thickness for both A(λ) and T(λ) maximum in our translucent OSC is in the range of 50nm to 80nm as shown in Figure 1. In this work, a 60 nm thick ITO-cathode contact was used as an optimal cathode layer, leading to a total transmittance 42% and an integrated absorptance of 34% for a translucent OSC of the type glass/ITO/PEDOT:PSS(40nm)/P3HT:PCBM(75nm)/Ca(10nm)/Ag(10nm)/ITO(60nm) (15).

Ex. 1016, Page 899

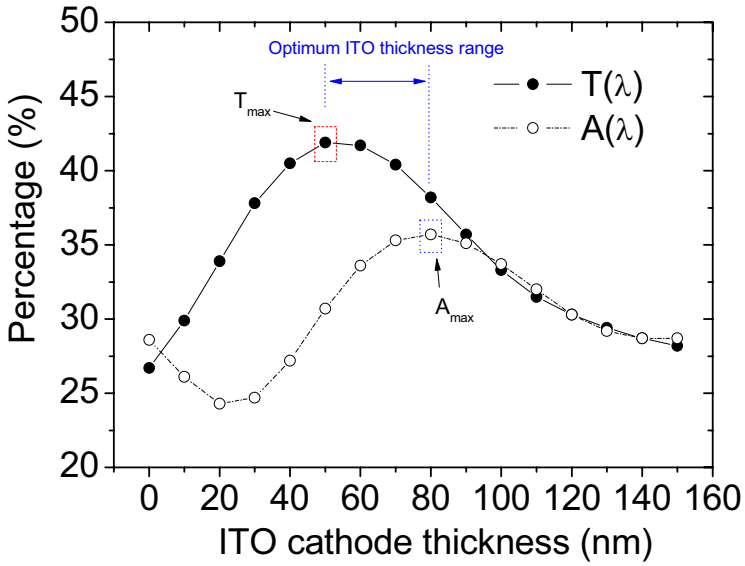


Figure 1: Calculation of the effect of the ITO cathode layer thickness on the integrated absorptance of P3HT:PCBM layer and the transmittance of translucent OSCs of the type: glass/ITO/PEDOT:PSS(40nm)/P3HT:PCBM(75nm)/Ca(10nm)/Ag(10nm)/ITO(0-160nm).

To evaluate the performance of our translucent OSCs, a control device with the configuration of glass/ITO/PEDOT(40nm)/P3HT:PCBM(75nm)/Ca(10nm)/Ag(100nm) (A1) was used as a reference device, and the I-V characteristics of the control OSC was compared with that of an structurally identical OSCs (A2 and A3) made with two different semitransparent cathodes of Ca(10nm)/Ag(10nm) and Ca(10nm)/Ag(10nm)/ ITO(60nm), respectively. The I-V characteristics of the control and translucent OSCs were measured under calibrated AM1.5 solar simulator in the nitrogen glove box and shown in Figure 2(a).

According to Figure 2(a), the control device A1 yielded a higher current density of 8.22 mA/cm2 in comparison with that of the translucent cells A2 (Jsc=5.20 mA/cm^2) and A3 (Jsc=6.89 mA/cm^2). At the same time, the fill factor of A2 is lower than that of A1 and A3 limiting the energy conversion efficiency. The increment in the photocurrent density of A2 in the forward bias region is slightly lower than that observed for the device A1 and A3 (Figure 2(a)). This indicated that the series resistance of A2 is higher than that of A1 and A3. The high series resistance is responsible for the low short circuit current density in A2 devices. For the comparison in Voc, A1 showed the highest value, followed by A3 and A2. Since the Voc is generally influenced by the work function difference between cathode and anode materials, the decrease in Voc for A2 and A3 could be attributed to the non-uniformity of thin Ca/Ag contact formated the polymer/cathode interface. A summary of the PV parameters measured for the devices A1, A2 and A3 is given in Table I.

The simulated and measured T(λ) of translucent OSCs made with two different cathode structures of Ca(10nm)/Ag(10nm) and Ca(10nm)/Ag(10nm)/ITO(60nm) are shown in Figure 2(b). It can be seen clearly that translucent OSCs with Ca(10nm)/Ag(10nm)/ITO(60nm) cathode have a higher T(λ) throughout the whole visible light wavelength region (red-solid) as compared to a structurally identical translucent OSC having a thin layer Ca(10nm)/Ag(10nm) cathode (red-dash). The

Ex. 1016, Page 900

Ca(10nm)/Ag(10nm) cathode in OSC causes a large amount of internal reflection at the metal/air interface, leading to a low T(λ). The simulated and measured T(λ) for both translucent OSCs with and without ITO index matching layer showed an overall good agreement with each other. The measured T(λ) of the translucent OSC does not overlap exactly as compared to that of the simulated T(λ) obtained for the same device, the deviation in the T(λ) between the experimental and calculated transparency in the devices may be attributed to the assumption of a perfect film smoothness, while it might not be true in reality as the non-homogeneity of the actual organic thin films may be expected.

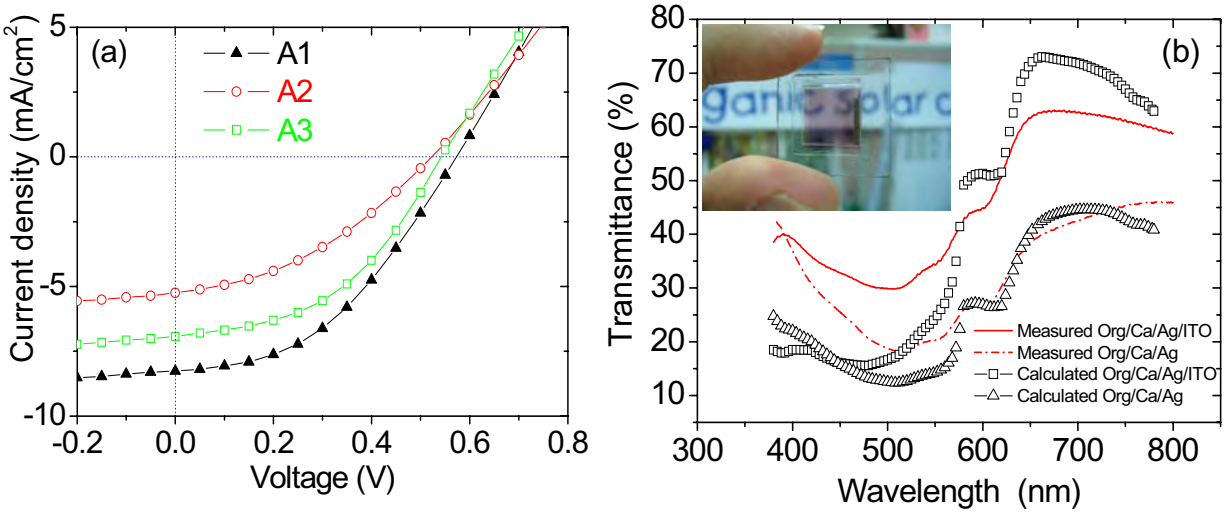


Figure 2: (a) A comparison of I–V characteristics measured for control cell A1, translucent solar cells A2 and A3. (b) Comparison of transmission spectral measured for translucent cells A2 (red-dash) and A3 (red-solid). The simulated curves for both configurations are drawn in open triangle and rectangle curve respectively. The inset is a photo picture showing a translucent polymer cell fabricated using an ITO-based cathode.

TABLE I. A summary of PV characteristics measured for control device (A1) and translucent OSCs A2 and A3.

Device #	V oc (V)	J sc (mA/cm^2)	FF (%)	PCE (%)
A1	0.57	8.22	43	2.02
A2	0.52	5.20	38	1.04
A3	0.54	6.89	46	1.70

With the results we have obtained so far, it suggests that the structure and the process optimization of ITO-based cathode are very important for the design and fabrication of high performance semitransparent OPV cells. The effect of the ITO cathode thickness on the total transmittance and integrated absorptance of P3HT:PCBM (75 nm) in semitransparent polymer PV cells was also analyzed. The results indicate that a maximum integrated transmission through semitransparent polymer PV cells can be achieved at an ITO layer thickness of 60 nm.

Ex. 1016, Page 901

III. Conclusion

The recent developments in organic solar cells with various translucent cathodes have been reviewed. We have developed a technique to improve the performance of translucent organic solar cells by optimizing the two competing indexes: power conversion efficiency and transmittance. We have also investigated the performances of P3HT/PCBM bulk heterojunction based solar cell by optimizing the light coupling efficiency of the translucent cathode, that is made up of Ca(10nm)/Ag(10nm)/ITO(60nm) multlayers. It was observed from our optical admittance analysis and the experimental results that the insertion of ITO layer as an optical index matching layer simultaneously enhances the PCE and T(λ) to 2% and 42% respectively. The high performance ITO-based cathode developed in this work can be formed at room temperature, without causing damages to the underlying functional organic layers, which is suited for application in both translucent and tandem solar cells. In addition, the ITO layer in the cathode also acts as a passivation layer to protect the underlying functional organic layers from any possible moisture attack from the environment in order to improve the lifetime of the solar cell. The translucent solar cells can be potentially used for applications in a variety of consumer electronics, including digital electronics, automotive, outdoor lifestyle and home appliances.

References

1. C. J. Brabec, N. S. Sariciftci, and J. C. Hummelen, Adv. Funct. Mater. 11, (2001) 15-26.
2. I. Riedel, J. Parisi, V. Dyakonov, L. Lutsen, D. Vanderzande, and J.C. Hummelen, Adv. Funct. Mater. 14, (2004) 38.
3. A. Yakimov, S. R. Forrest, Appl Phys. Lett. 80, 1667 (2002).
4. J. Xue, S. Uchida, B. P. Rand, S. R. Forrest, Appli. Phys. Lett. 85, 5757 (2004).
5. G. Dennler, H. J. Prall, R. Koeppe, M. Eggiger, R. Autengruber, and N. S. Sariciftci, Appl. Phys. Lett. 89, 073502 (2006).
6. A. Gadisa, K. Tvingstedt, S. Admassie, L. Lindell, X. Crispin, M. R. Andersson, W. R. Salaneck, and O. Inganäs, Synth. Met. 156, 1102 (2006).
7. V. Shrotriya, E. H. E. Wu, G. Li, Y. Yao, and Y. Yang, Appl. Phys. Lett. 88, 064104 (2006).
8. V. Shrotriya, E. H. E. Wu, G. Li, Y. Yao, and Y. Yang, Proc. of SPIE 6334, 633416 (2006).
9. A. Hadipour, B. de Boer, P. W. M. Blom, J. Appl. Phys. 102, 074506 (2007).
10. E. S. Zaus, S. Tedde, T. Rauch, J. Fürst, and G. H. Döhler, IEEE Trans Electron Devices 55, 2, (2008).
11. K. Kawano, N. Ito, T. Nishimori, J. Sakai, Proc. of SPIE 5938, 59381J (2005).
12. K. Kawano, N. Ito, T. Nishimori, and J. Sakai, Appl. Phys. Lett. 88, 073514 (2006).
13. T. Morimune, H. Kajii, and Y. Ohmori, Jpn. J. Appl. Phys., 44, 2815 (2005).
14. R. F. Bailey-Salzman, B. P. Rand, and S. R. Forrest, Appl. Phys. Lett. 88, 233502 (2006).
15. G. M. Ng, E. L. Kietzke, T. Kietzke, L. W. Tan, P. K. Liew, and F. R. Zhu, Appl. Phys. Lett. 90, 103505 (2007).

ECS Transactions, 18 (1) 875-880 (2009)
10.1149/1.3096549 © The Electrochemical Society

PV Efficiency Improved Large-area Dye-sensitized Solar Cells

H.Y. Yin, T.S. Zhang, X. Wang, L.F. Liu, Y. Wang, X.Y. Liu, J.F. Kang*
Institute of Microelectronics, Peking University, *E-mail: kangjf@pku.edu.cn

Several techniques including current-collecting metal fingers and double-layer TiO_2 electrode were applied to improve the conversion efficiency of large area dye-sensitized solar cells (DSSCs). Silver paste was used to fabricate the current-collecting metal fingers on the fluorine-doped tin oxide (FTO) glass substrate by screen-printing process. Then a single- or double-layer TiO_2 electrode was fabricated by using screen-printing process. The double-layer TiO_2 electrode consists of a transparent TiO_2 layer and a light-scattering TiO_2 layer. Results showed that applying the current-collecting metal fingers and double-layer TiO_2 electrode could effectively improve the conversion efficiency. The conversion efficiency of 5.89% was achieved in a DSSC module with an area of 50mm × 50mm by using the current-collecting metal finger cell of 45mm × 3.5mm and double-layer TiO_2 electrode under global air mass 1.5 condition.

I. INTRODUCTION

Dye-sensitized solar cells (DSSCs) are being widely studied as a promising solar cell technology because of its low-cost and environmental-friendship [1, 2]. Although conversion efficiencies exceeding 10% have been reported for small area DSSCs [2], it is still a problem to fabricate large area DSSCs with enough conversion efficiency for practical application [3, 4]. Inspired by the current-collecting metal fingers of commercial silicon solar cells, some researchers tried to apply the current-collecting metal fingers technique to the fabrication of the large area DSSCs to improve the conversion efficiency. The best aperture area efficiency of 5% was achieved [5]. Besides, it was reported that the double-layer TiO_2 electrode including a highly porous nano-crystalline TiO_2 layer and a light-scattering TiO_2 layer could enhance the conversion efficiency of DSSCs [6]

In this study, the impacts of metal fingers and double-layer TiO_2 electrode on the conversion efficiency of large size DSSC were investigated. Different structures of DSSC modules with/without metal fingers and single-/double-layer TiO_2 electrode were fabricated to compare the conversion efficiency. The results showed that photovoltaic performance of DSSCs was improved apparently by using both metal fingers and double-layer TiO_2 electrode techniques. The conversion efficiency of 5.89% was achieved in a DSSC module with an aperture area of 50mm × 50mm and an active area of 8×45mm×3.5mm under global air mass 1.5 condition.

II. EXPERIMENTAL

1. Materials

Commercially-available TiO_2 powders P25 (30 nm by Brunauer-Emmett-Teller (BET), 80% anatase and 20% rutile, via TiCl4-fumed gas synthesis, Degussa, Germany) and two kinds of screen-printable TiO_2 paste (DHS-TPP1, Dalian HeptaChroma SolarTech, Co., Ltd., China; T/SP, Solaronix SA Co., Switzerland) were used in the

Ex. 1016, Page 903

preparation procedure of screen-printable TiO_2 electrode. Ethyl celluloses powders (CP, #69010828, Sinopharm Chemical Reagent Co., Ltd.) were dissolved in ethanol solution at a concentration of 10 wt%. The bis (tetrabutylammonium) cis-di (thiocyanato) bis (2,20-bipyridine-4,40-carboxylic acid) ruthenium (II) (N719) (Solaronix SA Co., Switzerland) was used as a dye [7]. Conducting glass plates (Nippon Sheet Glass Co., Ltd., Tokyo, Japan), with fluorine doped tin oxide (FTO) film (sheet resistance 14Ω/□) were dipped into acetone and clean with ultrasonic washer for 30 minutes, rinsed with distilled water, and dried at 60℃. The liquid electrolyte used in this research was 0.3M LiI, 0.03M I2, and 0.5M ter-butyl-pyridine in CH_3CN.

2. Preparation of TiO_2 paste

The fabrication flow of TiO_2 paste using the P25 powders was described in Figure 1. The TiO_2 powders in ethanol were mixed with terpineol and a solution of ethyl cellulose in ethanol. After removing ethanol and water with a rotary-evaporator, the screen-printable paste consisting of 22.7% TiO_2 and 9% ethyl cellulose in terpineol was prepared.

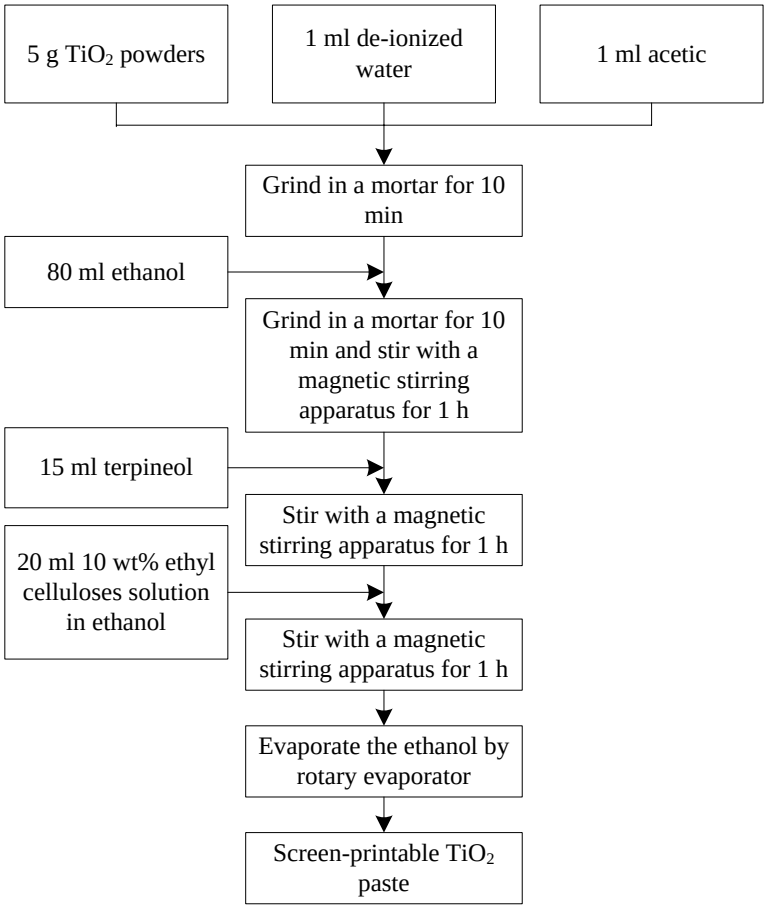


Figure 1. Fabrication flow of TiO_2 paste using the P25 TiO_2 powders;

3. TiO_2 electrode and current-collecting metal fingers fabricated by screen-printing

Stripe-type TiO_2 electrode and metal fingers were fabricated on 50mm×50mm FTO glass (Front Plate) by screen-printing. Three kinds of paste (P25, DHS-TPP1, T/SP) were applied in the study. The screen characteristics are as follows: material, polyester; mesh count, 300T mesh/inch; mesh opening, 51um; thread diameter, 34um; open surface, 36%; fabric thickness, 51um. (NSS300S, Smartmesh) After drying on a hot plate at 120℃for 30 min, electrodes were sintered at 450℃ for 3 hours. Current collecting metal fingers

Ex. 1016, Page 904

were also fabricated by screen-printing using silver paste (UNINWELL). The pattern of TiO_2 electrode and metal fingers were shown in Figure 2. Different-sized cells on the substrate were defined. In Figure 2(a), the 50mm×50mm module contains 4 cells with an area of 45mm×8mm; In Figure 2(b), the 50mm×50mm module contains 8 cells with an area of 45mm×3.5mm.

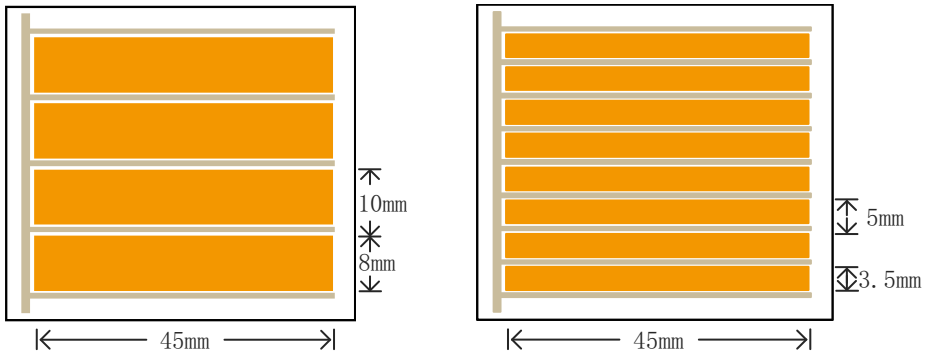


Figure 2. The patterns of TiO_2 electrode and metal fingers.

After drying on a hot plate at 100℃ for 1h, the front plate with TiO_2 electrode and metal fingers on FTO glass was immersed in a 0.5mM N719 dye solution at room temperature for 12h. The back plate with a Pt sputtered counter electrode was thermally pressed (100℃) onto the front plate with dyed TiO_2 film. The Surlyn film (25 mm) with two vents was used as the sealing spacer between the front and back plates. The electrolyte was introduced through the vents and the vents were sealed with Surlyn film. A schematic diagram of the cell construction is shown in Figure 3.

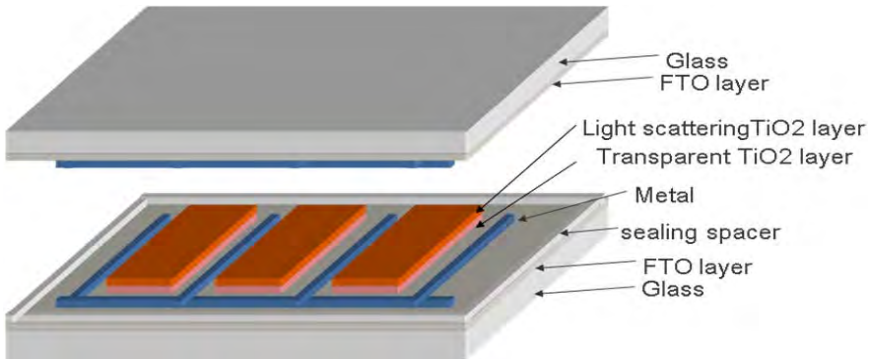


Figure 3. The schematic diagram of dye-sensitized solar cell module with metal fingers.

III. RESULTS AND DISCUSSION

1. Performances of DSSCs

Recent research showed that if the DSSCs scaled up to several square centimeters, the resistance of transparent conductive oxide (R_{TCO}) also increased, leading to low fill factor (FF) and low conversion efficiency [8]. To reduce the R_{TCO} of large area DSSCs, stripe-type TiO_2 electrode was fabricated with current-collecting metal (Ag) fingers on FTO glass. Figure 4 shows the photovoltaic characteristics of the DSSCs with and without

metal (Ag) fingers. The open-circuit photo voltage (V_{OC}), short-circuit photo current density (J_{SC}), fill factor (FF) and conversion efficiency (η) were summarized in Table I. The results indicate that, without metal fingers, the fill factor of DSSC module was only 25%, leading to a conversion efficiency of 0.76%. In contrast to this, with the introduction of metal fingers, the fill factors were increased to 57% and 63% in the DSSC module with the metal fingers cell of 45mm×8mm and 45mm×3.5mm, respectively. The corresponding conversion efficiency reached 4.31% and 5.49%.

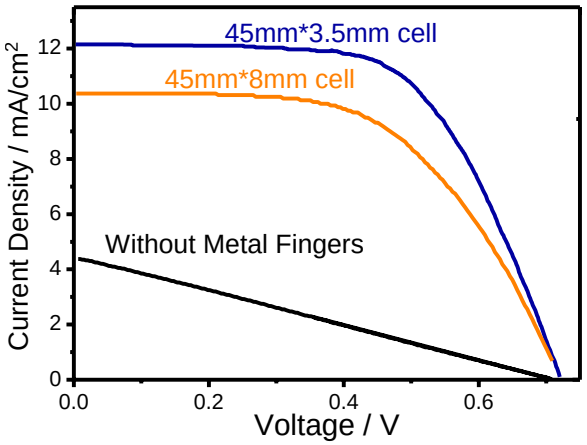


Figure 4. I-V characteristics of DSSC modules with and without metal fingers;

TABLE I. Photovoltaic performance of three DSSC modules with and without metal fingers

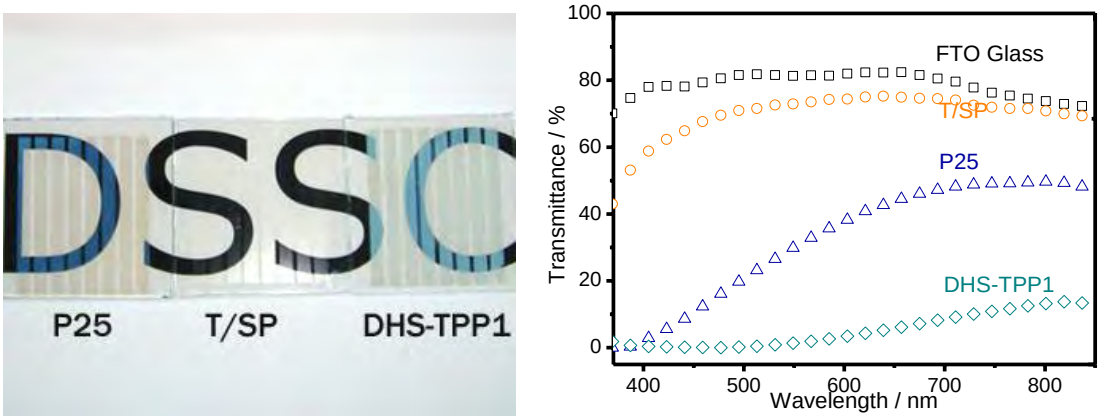
Type of module	Active area(cm^2)	Jsc(mA/cm^2)	Voc(V)	FF	η (%)
Without metal fingers	16	4.3	0.71	0.25	0.76
45mm*8mm cell	14.4	10.5	0.72	0.57	4.31
45mm*3.5mm cell	12.6	12.1	0.72	0.63	5.49

These results indicate that the introduction of metal fingers can significantly improve the conversion efficiency of large area DSSC module.

2. Single-layer and double-layer TiO_2 electrode

Figure 5(a) shows the image of screen-printed films with a background of black letters. The images show that Solaronix T/SP film is transparent and P25 and DHS-TPP1 films are semi-transparent. Figure 5(b) shows the transmittance spectra of the three screen-printed TiO_2 films in Figures 5(a). The T/SP layer transmitted 70% of incident light above 500 nm wavelength. The transmittance of P25 film increased linearly from 350 to 680 nm wavelength, and kept stable at approximately 50% from 680 to 850 nm. DHS-TPP1 film blocked the incident light below 550 nm wavelength and let the incident light slightly through above 550 nm.

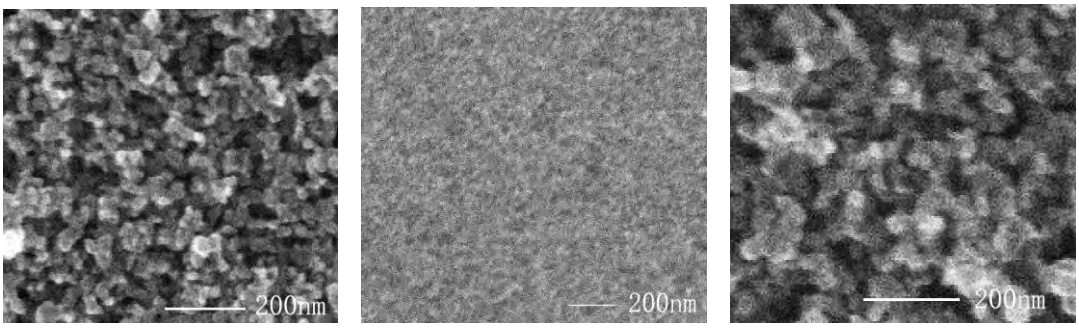
Ex. 1016, Page 906



(a) (b)

Figure 5. (a) Images of three kinds of screen-printed nanocrystalline-TiO_2 films with the background of black letters; (b) Transmittance spectra of the TiO_2 films corresponding to (a).

Figure 6 gives the SEM pictures of the surface morphology of the TiO_2 film prepared by the P25, Solaronix T/SP and DHS-TPP1 TiO_2 pastes. In Figure 6(a), both large-sized particles (rutile, 40 nm) and small-sized particles (anatase, 20 nm) can be observed. In Figure 6(b), the surface is smooth and uniform, which means the very small sized particles were formed. In Figure 6(c), large-sized particles (rutile, 40 nm) more than in Figure 6(a) is observed. It has been reported that large-sized particles can diffuse visible light effectively [9]. That's why the DHS-TPP1 film was able to transmit very little incident light (Figure 5), while the Solaronix T/SP film transmitted most of the incident light (70%). Here, the DHS-TPP1 film was used as a light scattering layer, while the Solaronix T/SP film was used as transparent layer.



(a) (b) (c)

Figure 6. SEM photographs of porous TiO_2 films showing the surface morphology of (a) P25, (b) Solaronix T/SP, (c) DHS-TPP1 TiO_2 paste;

Figure 7 shows the I-V characteristics of DSSC modules with the single layer (P25, DHS-TPP1, Solaronix T/SP) TiO_2 electrode and double layer (Solaronix T/SP+ DHS-TPP1) TiO_2 electrode. The open-circuit photo voltage (V_{OC}), short-circuit photo current density (J_{SC}), fill factor (FF) and conversion efficiency (η) were summarized in Table II. The DSSC modules with the single layer (P25, DHS-TPP1, Solaronix T/SP) TiO_2 electrode have the conversion efficiency of 5.59%, 5.49% and 2.5%. After introducing the "double-layer" system, the conversion efficiency of 5.89% was achieved. Thses results indicate that "double-layer" TiO_2 electrode can enhance the conversion efficiency.

Ex. 1016, Page 907

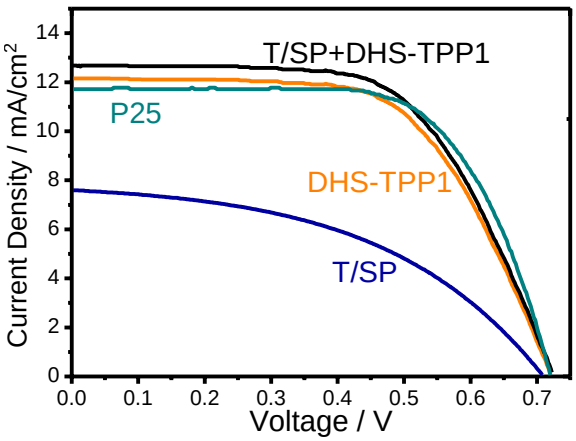


Figure 7. I-V characteristics of DSSC modules with single-layer TiO_2 electrodes fabricated by P25, DHS-TPP1, Solaronix T/SP TiO_2 paste and double-layer TiO_2 electrode fabricated by DHS-TPP1+ T/SP pastes;

TABLE II. Photovoltaic performance of three DSSC modules using different TiO2 paste

Type of module	Active area(cm^2)	Jsc(mA/cm^2)	Voc(V)	FF	η (%)
Solaronix T/SP	12.6	7.5	0.71	0.47	2.50
P25	12.6	11.6	0.72	0.67	5.59
DHS-TPP1	12.6	12.1	0.72	0.63	5.49
Solaronix T/SP + DHS-TPP1	12.6	12.8	0.72	0.64	5.89

IV. CONCLUTION

By using the current-collecting metal fingers and double-layer TiO_2 electrode techniques, the conversion efficiency of large-sized DSSC module were significantly increased. The conversion efficiency of 5.89% was achieved in a DSSC module with an aperture area of 50mm×50mm and an active area of 8×45mm×3.5mm under global air mass 1.5 condition. Our results suggest that the introduction of the optimized current-collecting metal fingers and double-layer TiO2 electrode structures may be a promising solution for the low cost, high performance, and large-sized DSSC module fabrication.

References

1. O'Regan B., Grätzel M., *Nature 1991*, **353**: 737–739.
2. Grätzel M., *Journal Photochemistry and Photobiology A: Chemistry 2004*, **164**: 23–14.
3. Okada K, et al., *Journal of Photochemistry and Photobiology A: Chemistry 2004*; **164**: 193–198.
4. Kawashima T, et al. *Journal of Photochemistry and Photobiology A: Chemistry 2004*; **164**: 199–202.
5. M. A. Green, et al., *Prog. Photovolt : Res. Appl.* **14**, 2006, pp. 45-51.
6. Seigo Ito, et al., *Thin Solid Films*, Vol. 516, May, 2008, pp. 4613–4619
7. A. Kay, M. Grätzel, *Chem. Mater.* 14 (2002) 2930–2935.
8. M.G. Kang, et al., *Sol. Energy Mater. Sol. Cells* 75 (2003) 475.
9. Seigo Ito, et al., *Progress in Photovoltaics: Research and Applications*, Vol. 15, pp. 603–612

ECS Transactions, 18 (1) 881-886 (2009)
10.1149/1.3096550 © The Electrochemical Society

Influence of Sulfurization Temperature on Structure and Semiconducting Properties of Sn_xS_y Thin Films

Cheng Shuying[a1], Peng Shaopeng[a], Yang Yongli[b]

[a] College of Physics and Information Engineering, Fuzhou University, Fujian Fuzhou, 350108, China
[b] Ministry of Education Key Laboratory of Analysis and Detection Technology for Food Safety, and Department of Chemistry, Fuzhou University, Fuzhou, Fujian ,350108,China

Sn_xS_y thin films were prepared by a two-step process. This process includes deposition of Sn films on glass substrates by thermal evaporation and sulfurization of Sn thin films in a vacuum condition at different temperatures of 150~300°C for 45min. The microstructures, composition and photoelectrical properties of the films were investigated. The results show that, there are some differences in structure, phase and morphology for the films synthesized at different sulfurization temperatures. All the films are of p-type conduction. With the increase of the sulfurization temperature, Hall mobility and carrier concentration of the films are increased, whereas the resistivity is reduced. The films sulfurized at 240°C are the best, and they are polycrystalline SnS films with orthorhombic structure. The SnS films have a strong {111} preferred orientation and good adhesion to the substrates. They are uniform with grain size of about 200~800 nm. The resistivity of the SnS film is 25.54Ω·cm.

1 Introduction

In recent years, thin films of tin sulfide（SnS）used as absorber layers in thin film solar cells, have attracted much attention for photovoltaic application. SnS belongs to orthorhombic structure. It has an optical gap of 1.30eV that is close to the optimal optical gap 1.50 eV of solar cells and a high absorption coefficient ($\alpha > 10^4 cm^{-1}$). In theory, the conversion efficiency of SnS film solar cells is up to 25%. SnS can be used as solar cell materials due to its low cost，easy solidification and being films. Furthermore, its elements Sn and S are abundant and non-toxic in nature. Thus, it will be potentially applied in the fabrication of photovoltaic devices (1).

At present, SnS films have been deposited by various methods (1-4). But SnS thin films prepared by a two-step process were seldom investigated, except for that K. T. Ramakrishna Reddy et al (5,6) reported on the synthesis of SnS films by this novel

[1] Corresponding author: Professor Cheng Shuying . Her research interest is in the field of solar cell materials and devices. Email: sycheng@fzu.edu.cn

two-step process and studied the structure and properties of SnS films. Therefore, here we also prepared SnS films by this process and investigated the influence of sulfurization temperature on structure and electrical properties of Sn_xS_y thin films so as to obtain SnS films with good semiconducting properties.

2 Experimental

Glasses (20mm*30mm*1 mm) were used as substrates in the experiment. The substrates were ultrasonically cleaned with acetone and ethanol in turn, rinsed with distilled water, and then dried by N_2 gas and a vacuum box prior to deposition.

The SnS thin films were prepared by vacuum thermal evaporation technique . The thermal evaporation system is DMDE-450 deposition equipment (made in China). To improve adhesion of SnS films to the substrates, the substrates were bombarded 15 min by ion beam before deposition. The substrates temperature was 50°C and the chamber was evacuated down to 3×10^{-3} Pa. The Sn films were deposited on the glass substrates, the thickness of the films was controlled by evaporation current and evaporation time. Their actual thickness is about 500~800nm measured by a profiler. The deposited Sn films were encased in a graphite case, and then was put into a vacuum furnace to be sulfurized. The sulfurization temperature and time were 150~300°C and 45min, respectively, and the vacuum degree was 6×10^{-1}Pa.

The structure and surface of the thin films were characterized by a Philips X′Pert-MPD X-ray diffraction spectroscopy（XRD） and a Philips XL30E Scanning Electron Microscope (SEM) with EDAX attachment, respectively . The dependence of the optical transmittance and reflectance on the wavelength was carried out with a Varian Cary 500 UV−VIS−NIR spectrophotometer. Based on Van der Pauw method, the electrical properties were determined by a HMS-3000 Hall measurement system.

3 Results and discussion

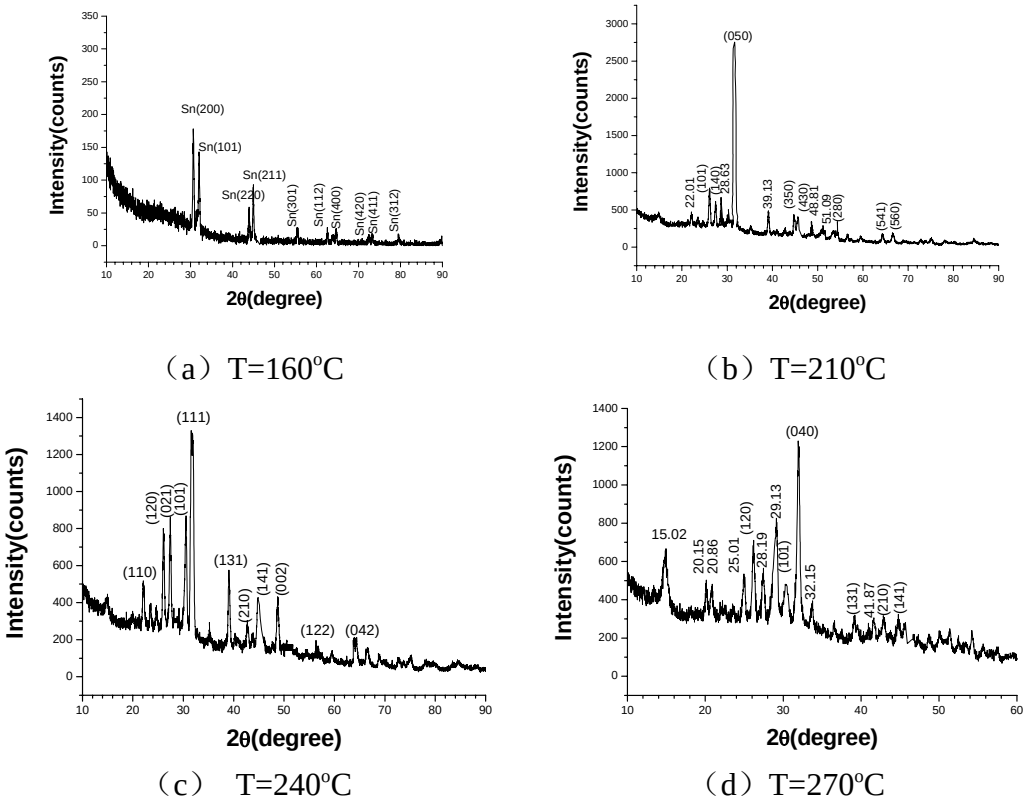
3.1 Structure and Surface

Fig.1 (a)~(d) show the XRD patterns of the films sulfurized at different temperatures, respectively. In Fig. 1(a), the three strong diffraction peaks (200)，(101)，(211) and other peaks, such as (220)，(301)，(112) and so on, are in good agreement with those of Sn (JCPDS04-0673). It indicates that the films are Sn at sulfurization temperature of 160°C. Maybe the S and Sn elements do not react due to too low sulfurization temperature.

In Fig. 1(b), peaks at 2θ equaling to 31.633°, 55.951° and 66.065° correspond to the strong diffraction planes (050), (280) and (560) of γ-Sn_2S_3 (JCPDS30-1379) and there are some other peaks of γ-Sn_2S_3 in the film. Simultaneously, the peaks appeared at 2θ=22.01° and 51.09° are respectively assigned to (110) and (211) planes of SnS, and the peak of 2θ=28.63° is assigned to (213) plane of S (JCPDS13-141). Furthermore, the peak of 2θ=39.13° is coincident with those from (131) plane of SnS and (301) plane of γ-Sn_2S_3. It represents that the primary composition of the films is γ-Sn_2S_3 at sulfurization temperature of 210°C , but maybe there is a little quantity of SnS and S.

All peaks in Fig. 1(c) agree very well with those of SnS (JCPDS39-354：

a=0.43291nm，b=1.11923nm，c=0.39838nm). The films prefer to grow along (111) plane. It indicates that the films are SnS with an orthorhombic structure at sulfurization temperature of 240°C.



（a）T=160°C　　（b）T=210°C

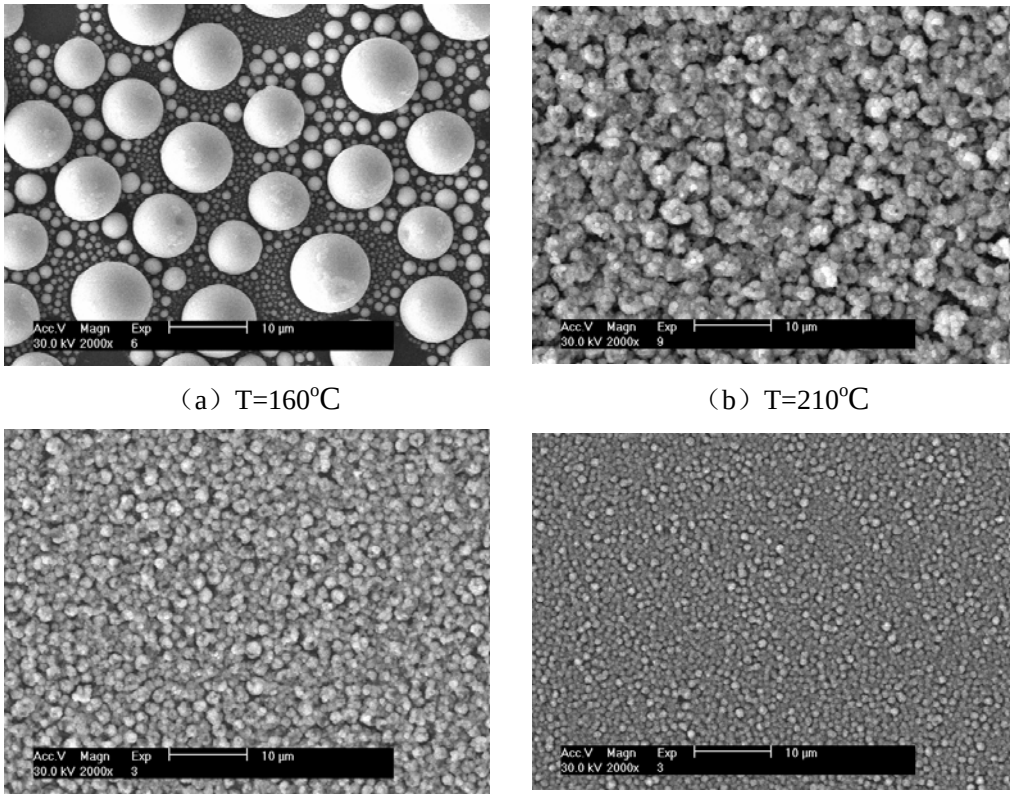
（c） T=240°C　　（d）T=270°C

Fig.1 XRD patterns of Sn_xS_y films synthesized at different temperatures

However, when the sulfurization temperature is 270°C, peak (101) of SnS becomes weaker, and (021) disappears. The preferred orientation of the films changes from (111) plane to (040) plane in Fig. 1(d). Probably because of high sulfurization temperature , enough energy is transmitted to the substrates, thus making the crystal growth orientation change. At the same time, the peaks at 2θ=15.02°, 28.19° and 32.15° are respectively assigned to the three strongest peaks (001), (100) and (101) of β-SnS_2 (JCPDS23-677) with hexagonal structure. It indicates that there exists a SnS_2 phase in the film. So it can be concluded that the material phase is varied with the rising of the sulfurization temperature .

Based on the above XRD profiles of Sn_xS_y films synthesized at different temperatures, the reaction mechanism may be as following: When the sulfurization temperature is too low, there is no enough energy to push the reaction between Sn and S elements in the films. When the sulfurization temperature is risen to 210°C, the main composition is Sn_2S_3 in the films. Orthorhombic SnS phase is dominant at 240°C. For sulfurization temperature at 270°C, SnS and SnS_2 coexist in the films. Therefore, the sulfurization temperature strongly affects the phases appeared in the films, which agrees with reference (5) though there are much differences in the sulfurization temperature. Since the difference between the control temperature and its real temperature is different for different annealing furnace. All the temperatures mentioned in the paper are the control temperatures which are a few tens less than the

real temperatures.



（a）T=160°C　　（b）T=210°C

（c）T=240°C　　（d）T=270°C

Fig.2. SEM pictures of Sn_xS_y films synthesized at different temperatures

Fig.2 shows SEM pictures of Sn_xS_y films synthesized at different temperatures. When the sulfurization temperature is 160°C, there are granular grains with much different sizes in the films. When the sulfurization temperature is risen to 210°C, the bigger grains gradually disappear and the size difference among grains become less though there are still some pinholes in the films and the films are not very uniform. When the sulfurization temperature is further risen to 240°C, the films become dense, smooth and uniform with good microstructure, their grain size is about 200~800nm. Therefore, the morphology of the films becomes better with the increasing of the sulfurization temperature .

For the films prepared on optimum parameters condition (sulfurization temperature of 240°C and sulfurization time of 45min), the EDAX analysis indicates that the atomic percent of Sn and S is : Sn-50.16at%，S-49.84at% . The Sn/S ratio was calculated to be 1.006/1, approaching 1.

3.2 Semiconductor properties

Hall measurement results show that all the films sulfurized at the range of 210～270°C are of p-type conductivity because the acceptors are provided by the vacancy of the Sn (7). Fig.3 shows the curve of bulk concentration and sulfurization temperatures. It can be seen that with the increase of the sulfurization temperatures, the carrier concentration of the films increases from 1.05×10^{16} cm^{-3} (undoped) to $1.43\times10^{17}cm^{-3}$. Because the rise of the sulfurization temperature reduces the strain or tensile stress,

Ex. 1016, Page 912

and the amount of dislocations and grain-boundary defects in the films, and it can improve crystallization of the films. In the mean time, more and more atoms in the unbalanced positions obtain energy, transfer to the relative balance positions, thus reducing the potential barrier between crystal lattices and making the crystal structure better .

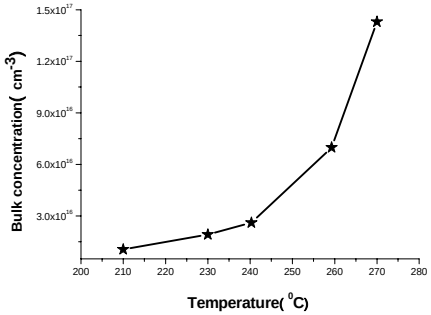


Fig.3 The curve of bulk concentration and sulfurization temperatures

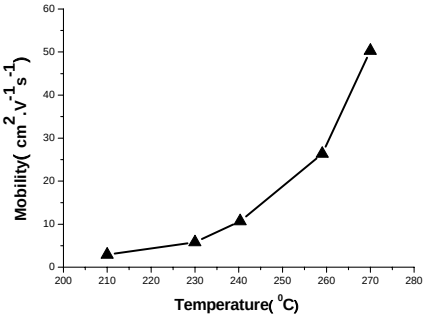


Fig.4 The curve of Hall mobility and sulfurization temperatures

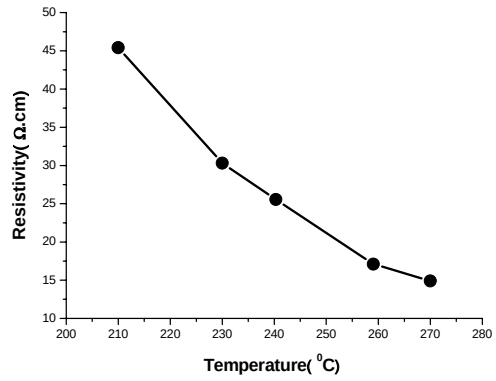


Fig.5 The curve of resistivity and sulfurization temperatures

With the increase of the sulfurization temperatures, the acceptor concentration of the films increases, namely, the vacancy of Sn increases. Because in the process of increasing the sulfurization temperature, heat vibration of atoms in all sites will intensify, and excess sulfur atoms will transfer to be interstitial atoms. Therefore, the concentration of acceptors will be increased.

Fig.4 shows the curve of Hall mobility and sulfurization temperatures. It can be seen that Hall mobility of the films increases from 2.92$cm^2 \cdot v^{-1} \cdot s^{-1}$ to 50.3 $cm^2 \cdot v^{-1} \cdot s^{-1}$ when the sulfurization temperature increases from 210°C to 270°C. The mobility is 10.7$cm^2 \cdot v^{-1} \cdot s^{-1}$ at sulfurization temperature of 240°C. Because at lower sulfurization temperature there are more defects and captured traps in the films , so the Hall mobility is lower. With the increasing of the sulfurization temperature, there is fewer defects and captured traps in the films, the morphology of the films is improved and the grains in the films get larger. Therefore, the carriers can move more freely in the crystal lattices.

Fig.5 shows the curve of resistivity and sulfurization temperatures. It can be seen that when the sulfurization temperature increases from 210°C to 270°C, the resistivity decreases from 45.62Ω·cm to 14.92Ω·cm. Because annealing effect makes the

resistivity decrease with the rise of sulfurization temperature. At sulfurization temperature of 210°C, the main composition of the films is Sn_2S_3, the resistivity is relatively high (8). At sulfurization temperature of 240°C, the resistivity is 25.54Ω·cm which is consistent with the order of the resistivity of SnS films prepared by spray method (2).

4. Conclusions

The Sn_xS_y films were prepared by a two-step process. The structure and electrical properties of Sn_xS_y films prepared at different temperatures were investigated. The research results show that: When the sulfurization temperature is too low, there is no enough energy to urge the reaction between Sn and S elements in the films. When the sulfurization temperature is risen to 210°C, the main composition is Sn_2S_3 in the films. Orthorhombic SnS phase is dominant at 240°C. For sulfurization temperature at 270°C, SnS and SnS_2 coexist in the films. The morphology of the films becomes better with sulfurization temperature increasing. Hall measurement shows that all the films are of p-type conduction, and with the increase of the sulfurization temperature, the Hall mobility and the carrier concentration of the films are increased, whereas the resistivity is reduced. The films sulfurized at 240°C are the best, and they are polycrystalline SnS films with orthorhombic structure. The SnS films have a strong (111) preferred orientation and good adhesion to the substrates. They are uniform and their grain size is about 200~800 nm. The resistivity of the films is 25.54Ω·cm.

Acknowledgment

The project-sponsored by SRF for ROCS, SEM (LXKQ0801) is gratefully acknowledged. The authors also wish to express their gratitude to funding from Fujian Provincial Department of Science & Technology and Fuzhou University, China (2008I0019 and XRC-0736).

References

1. Shuying Cheng, Yanqing Chen, Cicang Huang, et al. Thin Solid Films,**500**,96 (2006).
2. Koteawara Reddy, K. T. Ramakrishna Reddy. Solid-State Electronics, 49, 902 (2005)
3. P. Hamkare, A. V. Jadhav, P. A. Chate, et al. Alloy and Compounds, **463**, 581 (2008)
4. M.El-Nahass，H.M.Zeyada，M.S.Aziz，et al. Optical Materials，**20**, 159 (2002) .
5. K. T. Ramakrishna Reddy, P. Purandhara Reddy. Materials Letters, **56**, 108 (2002) .
6. K.T.Ramakrishna Reddy, P.Purandhara Reddy, P.K.Datta, et al. Thin Solid Films, **403/404**,116 (2002).
7. Ristov M，Sinadinovski G，Grozdanow I，et al. Thin Solid Films，**173**, 53 (1989) .
8. B.Thangaraju, P.Kaliannan. J. Phys. D: Appl. Phys.， **33**, 1054 (2000) .

Ex. 1016, Page 914

ECS Transactions, 18 (1) 887-894 (2009)
10.1149/1.3096551 © The Electrochemical Society

Thin Film InP Epitaxy on Si (001) Using Selective Aspect Ratio Trapping

J. Z. Li[a], J. Bai[a], J. M. Hydrick[a], J. G. Fiorenza[a], C. Major[a], M. Carroll[a], Z. Shellenbarger[b] and A. Lochtefeld[a]

[a] AmberWave Systems Corp., 13 Garabedian Drive, Salem, New Hampshire 03079
[b] Sarnoff Corporation, 201 Washington Rd, Princeton, NJ 08543

High quality InP thin films have been demonstrated in SiO_2 trenches on silicon via Aspect Ratio Trapping (ART), whereby defects arising from lattice mismatch (~8%) are trapped by laterally confining sidewalls. Double-buffer layers and two-step ART growth processes have been employed to trap vertical threading dislocations originating at InP/Si interface. InP film quality and optical properties have been analyzed using SEM, TEM and room temperature photoluminescence. Full trapping of dislocations has been demonstrated for trenches up to 400 nm in width without the additional formation of defects at the sidewalls above 500 nm initial growth. This approach shows great promise for the integration of III-V materials onto silicon.

Introduction

Integration of high quality III-V compounds on Si is a key technology for achieving optoelectronic integrated circuits (OEICs). For decades, there have been numerous attempts to realize InP/Si by various approaches both at wafer level and device level. Typical growth methods are graded buffer layer metalorganic chemical vapor deposition (MOCVD) (1), wafer bonding (2), hybrid growth of MOCVD with vapor mixing epitaxy (3) and epitaxial lateral overgrowth (4). Heteroepitaxy of high quality InP on Si poses well-known challenges, including (i) a considerable lattice constant mismatch (8%), much larger than that between GaAs and Si (4%), (ii) a large difference (50%) in thermal expansion coefficient and (iii) the polar/non-polar character of the two semiconductors. These result in a high density of dislocations, tetragonal distortion, antiphase domains and possible cracking in thin films (5, 6). At this time, despite demonstrations of InP-based laser diodes and resonant tunneling diodes on Si substrates (7-9), commercially available device quality InP/Si material is still lacking.

Very recently, we demonstrated substantially defect-free Ge and GaAs material grown in SiO_2 trenches on Si via the Aspect Ratio Trapping (ART) method (10,11), where threading dislocations arising from lattice mismatch are trapped at vertical sidewalls confining the growth region. In the current paper, we review the progress of InP growth on such an SiO_2-patterned Si substrate. High quality InP films were obtained in uncoalesced regions with a thin defect trapping region inside the SiO_2 trenches. Film quality improvement was achieved by using double-buffer layer growth, through optimizing a two-step ART process. The InP ART results are promising for achieving monolithic growth of III-V based optoelectronic devices on Si substrates.

Experimental

The substrates used in this study were 8" and 6" p-type on-axis Si (001). A thin SiO_2 layer was thermally grown on the Si substrate, followed by conventional photolithography and reactive ion etching techniques to form multi-section trenched patterns along the [110] direction. Each section has a uniform trench distribution, and the sections were divided by SiO_2 spacers. From section to section, the trench width and the distance between neighboring trenches vary from 0.18 to 0.35 μm and from 0.18 to 0.54 μm, respectively. Process details of SiO_2 mask patterning and substrate wafer cleaning have been previously described (10, 11). All processes were carried out using commercial Si processing facilities. After wafer preparation and prior to growth, the final trench height was about 720 nm. A low pressure MOCVD system with a vertical reactor was used for the growth. The In, P, Ga and As sources were trimethylindium, phosphine, trimethylgallium and arsine, respectively.

Initial experimentation was performed by using a two-step growth method and optimizing growth conditions on planar Si (001) substrates. Visual surface morphology inspection and X-ray diffraction measurement were used to roughly evaluate the film quality before conducting selective ART growth. In general, 1 μm InP grown on planar Si (001) substrates with full width at half-maximum (FWHM) of < 550 arc seconds (ω/2θ XRD) was regarded as an acceptable growth criteria. Early experiments indicated that using GaAs as a thin intermediate layer (InP/GaAs/Si) yields superior InP crystal quality to InP/InP/Si and InP/GaP/Si approaches. Therefore, we used the InP/GaAs/Si structure for InP growth via aspect ratio trapping throughout this study.

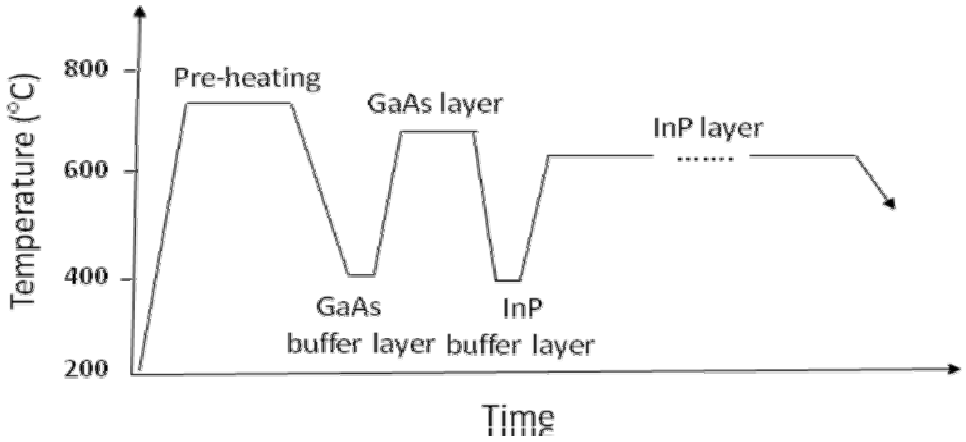


Fig. 1. Double buffer layer growth sequence of InP on Si substrate.

A typical sequence for growing InP on Si is shown schematically in Fig. 1. Si substrates are preheated at 750°C for 10 min in a H_2 atmosphere to remove thin oxide films prior to epitaxial growth. The first GaAs buffer layer is deposited at 400°C followed by a GaAs intermediate layer grown at 720°C. Then the growth temperature is decreased to 400°C for InP buffer layer growth, and finally the top InP layer is grown at 650°C. The growth rates are 5-7 nm/min for the buffer layers and 20-30 nm/min for the final InP layer. There is no thermal annealing step during the InP/GaAs/Si growth. Material quality has been analyzed via photoluminescence (PL), scanning electron

Ex. 1016, Page 916

microscopy (SEM) and transmission electron microscopy (TEM), and these results are compared to growth on unpatterned Si substrates.

Results and Discussion

InP ART growth investigations were initially conducted by exploring different binary compounds (e.g., InP, GaP, AlAs and GaAs) to serve as a buffer layer. Eventually, the InP/GaAs/Si growth sequence yielded the best results and was used for all further experimentation. Fig. 2 demonstrates typical results of trials carried out under different growth conditions, and illustrates patterned ART versus Si (001) substrate differences. Fig. 2(a) shows initial InP/Si growth using a 30 nm GaAs buffer layer. Though it has been reported that the GaAs buffer layer thickness can be as thin as a few tens of nanometers for growing InP on a blanket Si substrate (12), in this case, this buffer thickness was not sufficient to allow misfit dislocations to be trapped in the ART structure. Antiphase domains (APDs) and stacking faults were also observed from the Si surface to the upper InP layer.

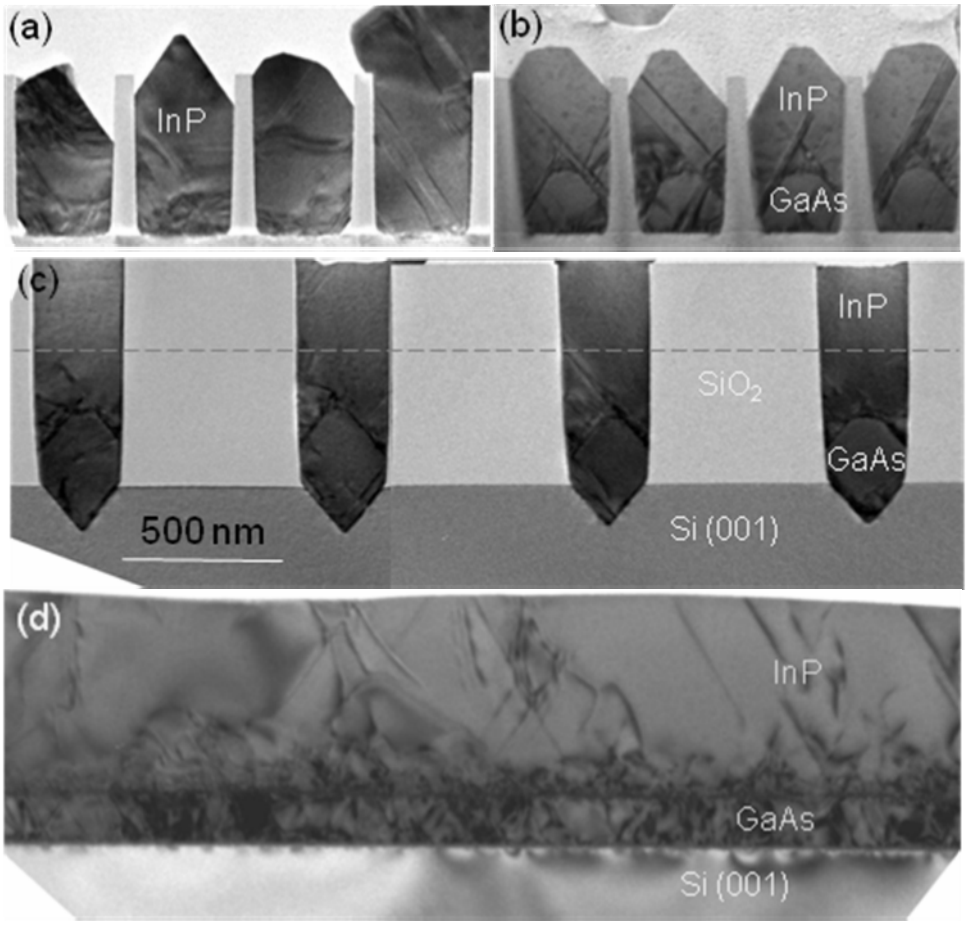


Fig. 2. Cross-sectional TEM images of (a) InP/GaAs/Si with a 30 nm GaAs buffer layer, (b) InP/GaAs/Si with 200 nm intermediate GaAs layer, (c) InP/GaAs/Si structure with v-grooved Si surface and modified growth conditions and (d) same growth as (c) but grown on a Si (001) substrate. All images are to the same scale.

Ex. 1016, Page 917

In Fig. 2(b), a 200 nm intermediate GaAs layer was added after the initial 30 nm GaAs buffer layer growth. This intermediate layer was thick enough to form a single crystal template. Many stacking faults and twins were observed to originate from the InP/GaAs interface along {111}. However, neither vertical threading dislocations nor APDs were observed. This implies that the single crystallized GaAs layer had played a role in blocking the APDs and misfit-related threading defects at the GaAs/Si interlayer, which agrees well with the previous GaAs/Si ART work (11).

Fig. 2(c) shows the results obtained by using a modified patterning approach and optimized growth conditions. Here, v-grooved surfaces at the trench bottom were formed by using KOH wet chemical etching. This v-grooved trench surface has distinct advantages for InP ART. First, KOH has large etch-selectivity among Si and SiO_2 surfaces (Si <100> : Si <111> : SiO_2 rates are 1000: 90 : 1 at room temperature) (13). Proper KOH etching can thoroughly remove contamination from the Si surface and create perfect, smooth {111} facets. Second, a v-grooved feature would effectively increase the desired aspect ratio in ART process and enable GaAs growth inside a pre-defined Si {111} enclosure, which can be expected to promote initial GaAs nucleation uniformity. In addition to the v-groove treatment, the InP layer growth conditions were modified as compared to the sample in Fig. 2(b). After the InP buffer growth, the first 100 nm of InP was deposited using a reduced growth rate at 600ºC, and then temperature was ramped to 650ºC for additional InP growth at the regular growth rate. Clearly, defects originating at the InP/GaAs interface were significantly reduced and eventually terminated at the SiO_2 sidewalls.

As shown in Fig. 2(c), crystallographic quality of the InP is high above the dotted line, where InP is essentially defect free, similar to the results shown previously for GaAs/Si growth by ART (11). This result clearly illustrates that the defects created by InP growth on a Si substrate have been effectively and significantly reduced by a two-step buffer, Aspect Ratio Trapping process. First, poplar/non-polar induced APD defects and 4% misfit defects have been effectively suppressed in the intermediate GaAs layer by optimized growth. Then, 4% lattice mismatched defects originating at the InP/GaAs interface were reduced via ART in a region above the intermediate GaAs layer. It should noted that the top of the APD-free InP film in Fig. 2(c) (which was originally slightly overgrown, above the SiO_2 surface) was unintentionally removed during TEM sample preparation.

For comparison, cross-sectional TEM of InP grown on Si (001) substrate, from the same growth as indicated in Fig. 2(c), is shown in Fig. 2(d). For the blanket film on Si, various mixed defects exist with high threading dislocation defect density in the InP layer.

Fig. 3 shows the room temperature PL measured from samples shown in Fig. 2(c) and (d). The PL peak intensity of InP grown on an ART patterned substrate is about 1/3 as strong as that of InP/InP, with identical FWHM, 27 nm. In contrast, the PL intensity of InP on Si is about 1/60 of that of InP/InP sample with broader FWHM, 38 nm. A slight blue shift shown in spectrum (a) and (c) is probably because the InP on the GaAs intermediate layer is slightly compressed.

Ex. 1016, Page 918

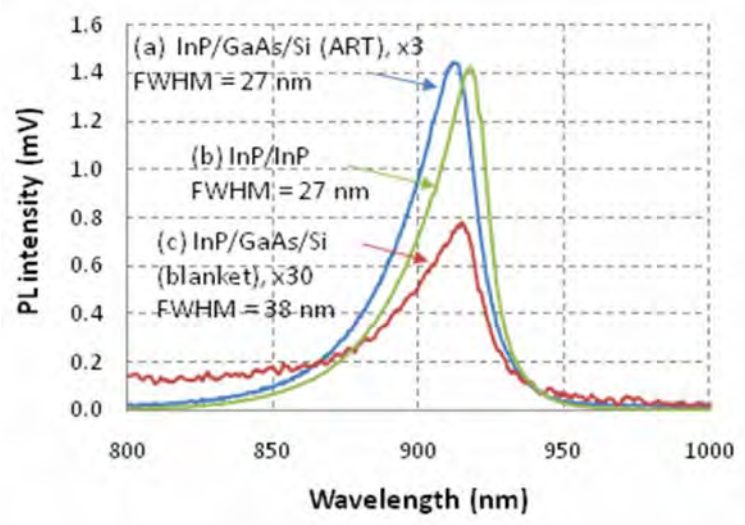


Fig. 3. Room temperature photoluminescence spectra of 0.8 μm InP grown on (a) SiO_2 trench patterned Si (001) in 250 nm trench / 250 nm spacing section, (b) InP substrate and (c) Si (001) substrate. These three samples were grown under the same growth conditions as described in Fig. 2 (c).

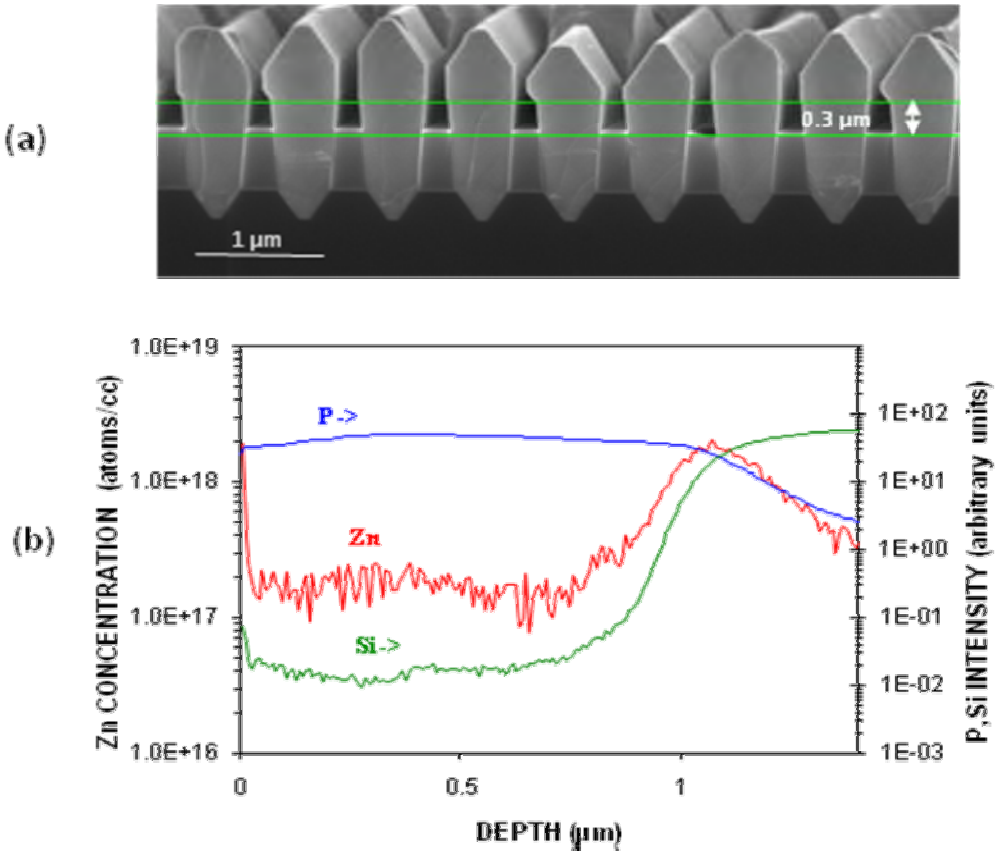


Fig. 4. (a) Cross-sectional SEM image of 1.5 μm p-InP grown on SiO_2-patterned Si (trench/space = 2:1 section) for SIMS analysis. 0.3 μm of SiO_2 has been removed via HF. (b) SIMS profile measured from sample (a).

Impurity contamination to the InP layers grown inside the trenches has been examined through second ion mass spectroscopy (SIMS) analysis. Fig. 4(a) shows the SIMS sample, a 1.5 μm uncoalesced InP film grown under the same conditions as illustrated in Fig. 2(c), with a trench/space ratio of 2:1. This sample was intentionally zinc-doped with a target doping level of $2x10^{17}$ cm^{-3}. To exclude sputtering-induced Si and O "contamination" originating from the SiO_2 mask layer, part of the SiO_2 mask layer was removed by selective chemical etching. The average height of the InP stripes above the remaining SiO_2 was about 0.8 μm. Fig. 4(b) shows a background Si level in upper InP layer of about $3\text{-}4x10^{16}$ cm^{-3}, which is an acceptable level for building optoelectronic devices. Lower background Si contamination could be realized by further optimizing the V/III ratio and growth temperature during upper InP epitaxy without altering the ART characteristics.

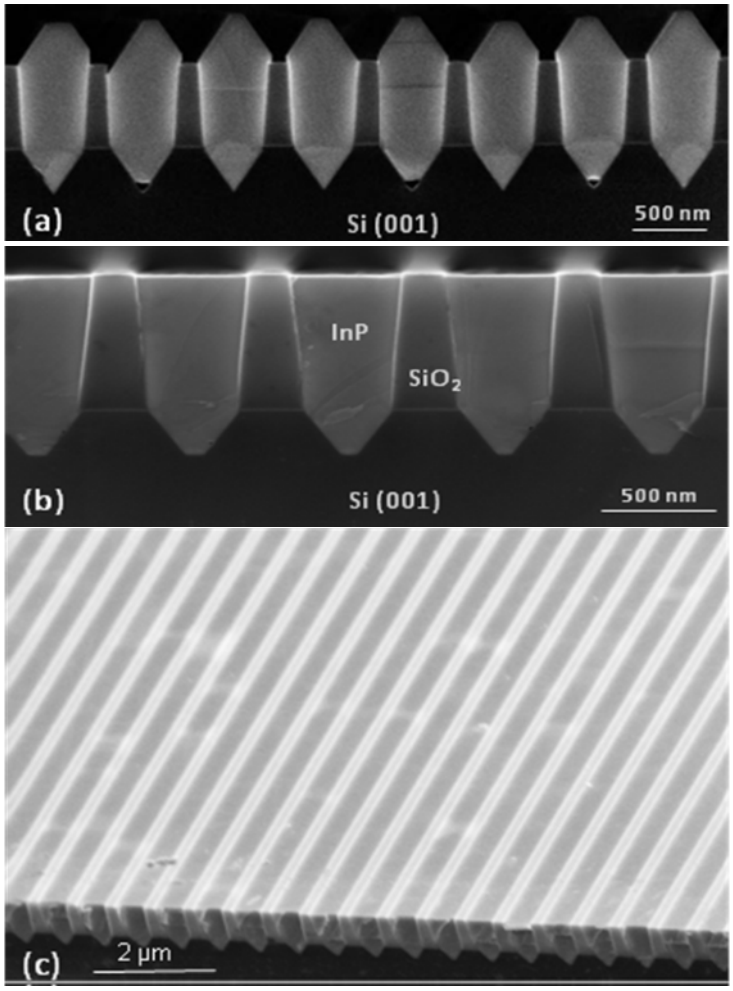


Fig. 5. Cross-sectional SEM images of (a) 1 μm p-InP grown on SiO_2-patterned Si (trench/space = 2:1 section) and (b) InP/GaAs/Si after chemical mechanical polishing (CMP). (c) Top view SEM images of InP/GaAs/Si after CMP.

To construct InP-based optoelectronic devices, particularly transport electronic devices, on the uncoalesced submicron films described above, a planar InP template surface is required. Fig. 5 illustrates results achieved by using chemical mechanical polishing (CMP), performed on a Logitech Orbis CMP. Fig. 5(a) shows the cross-sectional SEM

Ex. 1016, Page 920

image of a typical uncoalesced InP/GaAs/Si structure grown using the ART method, as described above. The top surface features have a combination of {111} and {100} facets. Cross-sectional and top view SEM images after CMP are shown in Figs. 5(b) and (c), respectively. As expected, low defect, APD-free InP stripes with flat surface was obtained. The surface width of the planar InP stripe is about 500 nm, and the stripes are separated by 200 nm SiO_2, with total InP film thickness less than 1 μm. Promisingly, the planar surface area of these one-dimensional InP stripes would be sufficient templates for fabricating InP-based heterojunction bipolar transistors, resonant tunneling diodes and other optoelectronic devices.

Conclusion

High quality, uncoalesced InP thin films have been grown by MOCVD in SiO_2 trenches on Si (001) via the Aspect Ratio Trapping method. Dislocation defects, stacking faults and twinning defects were trapped by laterally confining sidewalls through an optimal two-step ART process. Experimental investigation indicated that the addition of a v-grooved Si surface to the ART process can more effectively trap misfit defects and APDs at a GaAs/Si intermediate interface. Stacking faults and threading dislocations originating at the InP/GaAs interface can be suppressed by optimizing ART growth conditions, without involving thermal cycle annealing. SIMS analysis shows a negligible low level of Si diffusion in the InP layer. A mirror-like flat surface has been created by chemical mechanical polishing with a template layer thickness of less than 1μm. This approach provides great promise for realizing monolithic growth integration of III-V based optoelectronic devices with adapted Si technology.

Acknowledgments

The authors would like to acknowledge Logitech and SVTC for assistance with aspects of this work.

References

1. N. J. Quitoriano and E. Fitzgerald, *J. Appl. Phys*. **102**, 033511 (2007).
2. H. Park, A. W. Fang, S. Kodama and J. E. Bowers, *Optics Express*, **13**, 9460 (2005).
3. M. Tachikawa, T. Yamada, T. Sasaki, H. Mori and Y. Kadota, *Jpn. J. Appl. Phys.,* Part 2 **34**, L657 (1995)
4. S. Lourdudoss, E. R. Messmer, O. Kjebon and G. Landgren, J. Crystal Growth, **152** 105-114 (1995).
5. P. Pirouz, F. Ernst and T.T. Cheng, *Heteroepitaxy on Silicon*, *Mater. Res. Soc. Proc.* **116**, Pittsburg, PA (1988).
6. M. Sugo, M. Yamaguchi, and M.M. Al-Jassim, *J. Crystal Growth*, **99** 365 (1990)
7. T. Yamada, M. Tachikawa, T. Sasaki, H. Mori and Y. Kadota, *Appl. Phys. Lett*. **70** 1614 (1997).
8. T. Makimoto, K. Kurishima, T. Kobayyashi and T. Ishibashi, *Jpn. J. Appl. Phys.,* **30**, 3815 (1991).
9. W. Prost, V. Khorenko, A. C. Mofor, A. Bakin, E. Khorenko, S. Ehrich, H.-H. Wehmann, A. Schlachetzki and F. –J. Tegude, *Proceedings of ESSDERC*, Grenoble, France, Paer 3. B. 3 (2005)

10. J.-S. Park, J. Bai, M. Curtin, B. Adekore, M. Carroll, and A. Lochtefeld, *Appl. Phys. Lett.* **90**, 052113 (2007).
11. J. Z. Li, J. Bai, J.-S. Park, B. Adekore, K. Fox, M. Carroll, A. Lochtefeld, and Z. Shellenbarger, *Appl. Phys. Lett*. **91**, 021114 (2007)
12. Y. Kohama, Y. Kadota and Y. Ohmachi, *Jpn. J. Appl. Phys.*, **28**, 1337 (1989).
13. H. Seidel, L. Csepregi, A. Heuberger and H. Baumgartel, J. Electrchem. Soc., 137, 3612 (1990).

ECS Transactions, 18 (1) 895-900 (2009)
10.1149/1.3096552 © The Electrochemical Society

Sub-micrometer Organic Field Effect Transistors

L. Shang, M. Liu, Z. Ji, G. Liu, X. Liu, J. Liu, and H. Wang

Key Laboratory of Nanofabrication and Novel Devices Integration Technology, Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China

Pentacene organic field effect transistors with a series of sub-micrometer channel have been fabricated and characterized. Source and drain metal electrodes was made by electron-beam lithography and lift-off process. Image reversal photo resist was used to make integrated mask to pattern the organic semiconductor film to reduce the off-state current. And thiol treatment was applied on gold electrodes to modify their surface energy and improve the charge carrier injection. All the devices can behavior as typical p-type transistor, and have obvious saturation. Using electrode and dielectric surface treatment, device performances have significant improvement. All of the performances, including mobility, threshold voltage and on/off current ratio will change with the channel length decreasing because of the low quality interface between the OSC and dielectrics.

Introduction

Organic field-effect transistors (OFETs) have attracted much attention and enormous advances have been made in recent years as a result of worldwide activity in numerous research groups. The potential applications of OFETs include large-area display, sensors and electronic barcodes etc (1). However, most of the reported devices have channel longer than one micrometer. Few of reports demonstrated the characteristics of OFETs with sub-micrometer channel (2-14), which is necessary for high frequency applications. In this work, the device performances *vs.* channel length were studied systemically based on a series of sub-micrometer pentacene OFETs. Integrated resist mask was used to pattern the organic semiconductor (OSC) film, and thiol treatment was employed to improve the interface between the metal electrodes and OSC.

Experiment

Standard planar OFETs using the bottom contact geometry (Fig.1a) have been fabricated using EBL. Heavily doped silicon serve as common gate electrode and supporting substrate. A 60 nm thick thermally grown SiO_2 was used as dielectric layer. On this oxide layer, the source and drain electrodes, consisting of 20 nm thick Au with 2 nm Cr adhesion layer, were defined by EBL, using JEOL JBX 5000 system with a PMMA/copolymer bilayer resist. Interdigital patterns have been applied for metal electrodes. All the fingers are 0.1 μm wide. 9 kinds of channel length (0.1-0.9 μm, with a step of 0.1 μm) have been employed. All the devices have the same length /wide ratio of 1/100. Before the deposition of OSC, the samples were either used to form integrated

Ex. 1016, Page 923

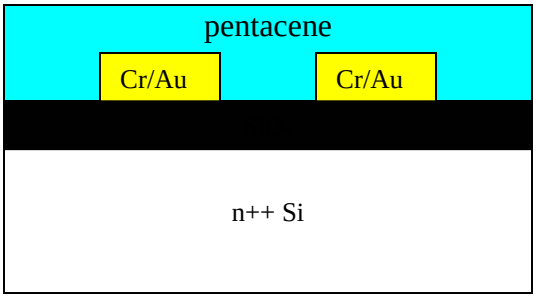


Figure 1. Schematic of bottom contact OFET

masks, or treated by thiol solution. Patterning active layer can reduce the leakage of OFET efficiently. In this work, an image reversal photo resist (Clariant AZ5214) was used to make integrated mask because it can easily form undercut. 1-Octadecanethiol (ODT, $CH_3(CH_2)_{17}SH$) was applied to the Au electrodes to modify the surface energy to improve the charge injection. The samples with gold electrodes were immersed in 1 mMol/L ethanolic thiol solutions for about 24 h followed by careful rinsing with clean ethanol and drying with pure N_2. Pentacene (Aldrich) was deposited onto the wafers using a vacuum evaporator (background pressure ~4×10^{-4} Pa) after a gas phase OTS treatment. During the deposition, the wafers were kept at 60 °C to obtain a high order film. The deposition speed and final thickness were 0.5Å/s and 50 nm, respectively.

The electrical characteristics of the devices were measured in air at room temperature using a Keithley 4200 semiconductor characterization system in company with a CASCADE RF1 probe station. Hole mobilities (μ) were extracted from square-root plots of drain current (I_{DS}) vs. gate voltage (V_{GS}), through the formula of saturation regime,

$$I_{DS}=WC_i(V_{GS}-V_T)^2/2L$$

where C_i is the capacitance of the gate dielectric per unit area (57.5 nF/cm^2) and V_T is the threshold voltage.

Result and Discussion

Figure 2 shows the output and transfer characteristics of OFETs with integrated mask to pattern the active layer. For simplicity, only the characteristics of three kinds of channel length devices have been given out. The output curves indicated that all the devices, including other 6 kinds of channel length, can behavior like a p-type filed effect transistor. All kinds of devices exhibited obvious gating effect, but the effect become weak as the channel length decreasing due to short channel effect. However, the saturation section still exists, which proved that super-linear is not an inherent limitation of sub-micrometer OFETs (8). The field-effect mobilities were extracted from the square root of drain current *vs.* gate voltage curves. The values are only about 10^{-5} cm^2/Vs, which are much smaller than that of long channel devices. Using the same materials and OSC depositing procedure, mobility of 0.5 cm^2/Vs can be achieved in 40 μm channel top-contact devices. Therefore, the severe performance degradation of these devices is attributed to the device configuration and the additional fabrication processes. In long

Ex. 1016, Page 924

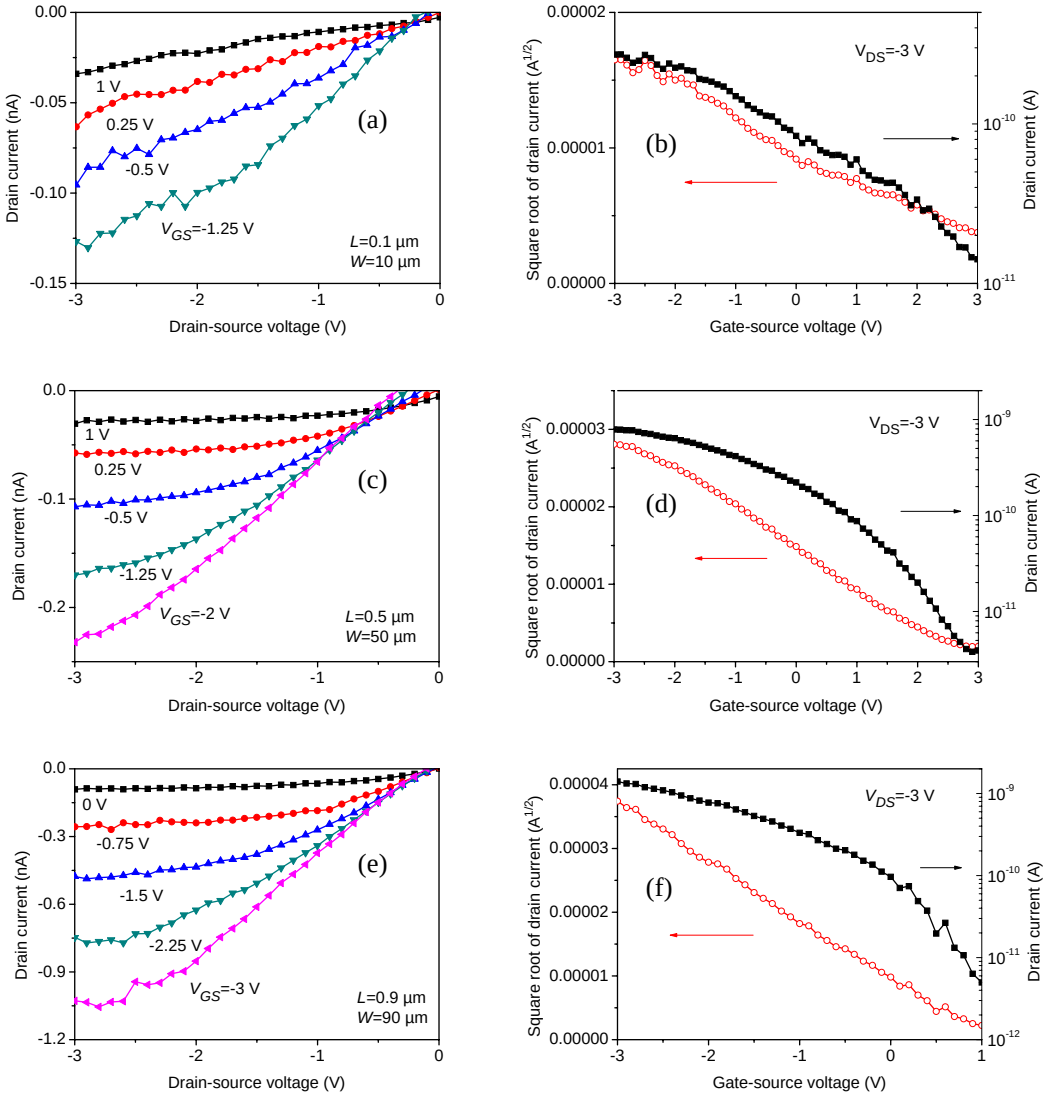


Figure 2. Electronic characteristics of different channel OFETs with active layer patterned. (a), (c) and (e) are the output characteristic curves of 0.1, 0.5 and 0.9 μm channel device respectively. (b), (d) and (f) are the transfer characteristic curves of 0.1, 0.5 and 0.9 μm channel device respectively.

channel bottom-contact devices, the large grain, well ordered OSC film in the center of the channel will change into a microcrystalline structure near the edge of the channel. This crystalline structure at this edge causes the performance degradation of the bottom contact devices compared with top-contact devices. In sub-micrometer channel device, much more film areas turn to be microcrystalline because the short distance among metal electrodes. The integrated mask fabrication process induced impurities attached onto the electrodes and the dielectrics, decreasing the interface quality which is important for device performance. Thus, it is necessary to modify the surface of channel and electrode before deposition of pentacene to improve the device performance.

In this work, ODT treatment was applied to modify the surface of Au electrodes and O_2 plasma was used to clear the samples surface. Figure 3 gives out the characteristics of

three kinds of channel length devices modified using these methods. The corresponding on-state currents have 2 orders of magnitude increase, which indicate the carrier injection have been improved by these treatments. And these devices can work at a higher voltage. Figure 4 show the performance variation *vs.* channel length compared with those of the integrated mask patterned devices. The mobilities and ratios of on/off current have been

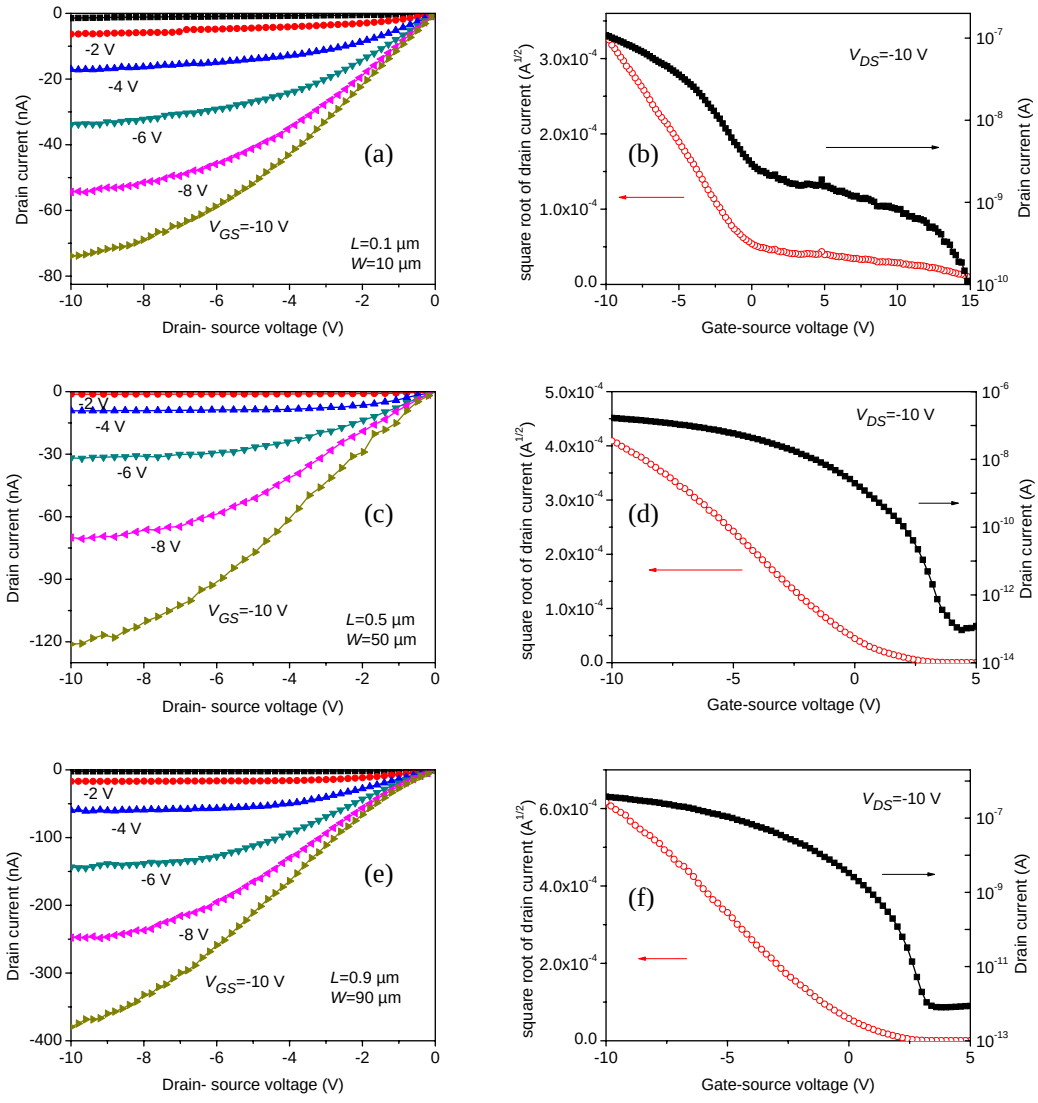


Figure 3. Electronic characteristics of different channel OFETs based on ODT treated Au electrodes. (a), (c) and (e) are the output characteristic curves of 0.1, 0.5 and 0.9 µm channel device respectively. (b), (d) and (f) are the transfer characteristic curves of 0.1, 0.5 and 0.9 µm channel device respectively.

improved after using the modification. The mobilities of modified devices can reach 10^{-3} cm^2/Vs, which is similar to the reported results in (12) using the same material. All of the performances, including mobility, threshold voltage and on/off current ratio will change with the channel length decreasing. The mobilities have several times decrease when the channels descend from 0.9 to 0.1µm (figure 4(a)). Devices with channel longer than 0.2 µm have positive threshold voltages (<2 V), however, the 0.1 µm transistors have

negative V_T (figure 4 (b)). Most of the devices have on/off ratios high as 10^6, but the ratio drops to 10^3 for 0.1μm devices (figure 4(c)). This big drop is because of the low quality interface between the OSC and dielectrics results in decrease of on-state current and increase of off-state current. The thermal evaporated organic semiconductors were deposited both on the metal surface and the dielectric surface. At the edges of electrodes, the OSC form a lot of strip-liked grains. These grains will contact both of source electrodes and drain electrodes when the channel is near 300 nm, and the channel become

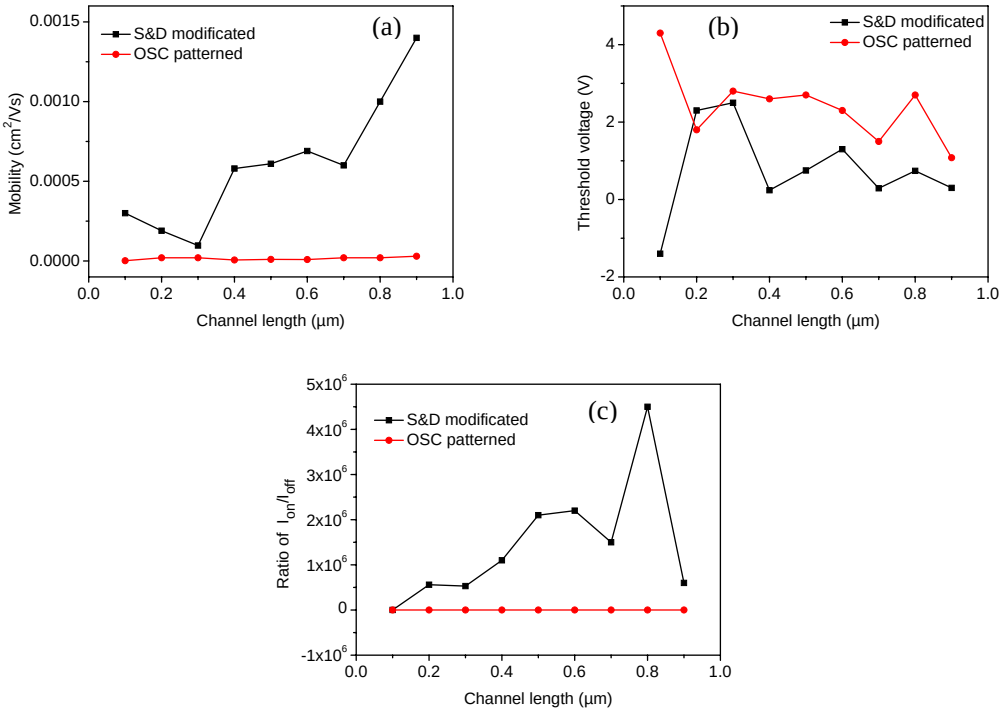


Figure 4. different device characteristics *vs.* channel length. (a) mobility, (b) threshold voltage, and (c) ratio of I_{on}/I_{off}.

shorter, the more grains will contact two electrodes. This kind of grains will block the pentacene particles from reaching the dielectric surface and produce a lot of gaps between the active layer and insulator layer. In the gap region, the gate can not control the carrier accumulation in active layer and lose the capability of modulating current in the channel. This is the reason why the ratios of on/off current and mobilities decline to very small in short channel length.

Conclusion

This work systemically studied the characteristics of sub-micrometer organic field effect transistors. Standard planar bottom-contact OFETs have been fabricated using electronic beam lithography. Pentacene is used as the semiconductor materials. Image reversal photo resist was used to make integrated mask to pattern the organic semiconductor film to reduce the off-state current. And thiol treatment was applied on gold electrodes to modify their surface energy and improve the charge carrier injection. The mobilities have several times decrease when the channels descend from 0.9 to 0.1μm (figure 4(a)).

Ex. 1016, Page 927

Devices with channel longer than 0.2 μm have positive threshold voltages (<2 V), however, the 0.1 μm transistors have negative V_T (figure 4 (b)). Most of the devices have on/off ratios high as 10^6, but the ratio drops to 10^3 for 0.1μm devices (figure 4(c)). This is because of the low quality interface between the OSC and dielectrics due to the short distance among the electrodes.

Acknowledgments

This work was partially supported by China National 973 Program (2006CB806204), and National Natural Science Foundation of China under Grant No. 90607022, 60676001, 60676008.

References

1. A. Dodabalapur, L. Torsi, and H. E. Katz, *Science*, **268**, 270 (1995)
2. E. T. Jones, O. M. Chyan, and M. S. Wrighton, *J. Am. Chem. Soc.*, **109**, 5526 (1987)
3. S. Franssila. J. Paloheimo and P. Kuivalainen, *IEEE Electronics Letters*, **29**, 713 (1993)
4. J. A. Rogers, A. Dodabalapur, Z. Bao, and H. E. Katz, *Appl. Phys. Lett.*, **75**, 1010 (1999)
5. E. L. Granstrom and C. D. Frisbie, *J. Phys. Chem. B*, **104**, 12202 (2000)
6. J. Collet, O. Tharaud, A. Chapoton, and D. Vuillaume, *Appl. Phys. Lett.*, **76**, 1941 (2000)
7. M. D. Austin, S. Y. Chou, *Appl. Phys. Lett.*, **81**, 4431 (2002)
8. Y. Zhang, J.R. Petta, S. Ambily, Y. Shen, D. C. Ralph and G. G. Malliaras, *Adv. Mater.*, **15**, 1632 (2003)
9. J. Zaumseil, T. Someya, Z. Bao, Y. Loo, R. Cirelli, and J. A. Rogers, *Appl. Phys. Lett.*, **82**, 793 (2003)
10. N. Stutzmann, R. H. Friend, and H. Sirringhaus, *Science*, **299**, 1881 (2003)
11. A.P. Kam, J. Seekamp, V. Solovyev, C. Clavijo Cedeno, A. Goldschmidt, C.M. S. Torres, *Microelectronic Engineering*, **73**, 809 (2004)
12. G. Leising, B. Stadlober, U. Haas, A. Haase, C. Palfinger, H. Gold, and G. Jakopic, *Microelectronic Engineering*, **83**, 831 (2006)
13. T. Jung, B. Yoo, L. Wang, and A. Dodabalapur, B. A. Jones, A. Facchetti, M. R. Wasielewski, and T. J. Marks, *Appl. Phys. Lett.*, **88**, 183102 (2006)
14. L. Shang, M. Liu, D. Tu, L. Zhen, G.Liu, and X. Liu, *Mater. Res. Soc. Symp. Proc.*, **1091**, AA07-80 (2008)

ECS Transactions, 18 (1) 901-906 (2009)
10.1149/1.3096553 © The Electrochemical Society

Resistive Switching Characteristics of MnO_x-based ReRAM

Sen Zhang[a], Shibing Long[a], Qi Liu[a, b], Qin Wang[a], and Ming Liu[a]

[a] Laboratory of Nano-Fabrication and Novel Device Integrated Technology, Institute of Microelectronics,Chinese Academy of Sciences, Beijing, China
[b] College of Electronics and Technology, Anhui University, Hefei, China

Thin film of MnO_x was fabricated in the form of Pt/MnO_x/Pt and Pt/MnO_x/Al and both devices exhibited reversible resistive switching behaviors under sweeping voltage. The resistive switching was believed to be due to the formation and rupture of conductive filaments. The device of Pt/MnO_x/Al showed a better endurance performance than Pt/MnO_x/Pt, its high/low resistance ratio was over three orders and the state retention was reliable. All of the results suggested that the Pt/MnO_x/Al has a potentiality for nonvolatile resistance memory application.

Resistance random access memory (ReRAM) is a competitive candidate for the next generation nonvolatile memories, it is based on the electric field induced resistive-switching property. Various materials have been found to exhibit resistive switching phenomenon including ferromagnetic materials (1), doped perovskite oxides (2), organic materials (3) and transition metal oxides (4). Among them, transition metal oxides are extensively studied recently due to their good resistive-switching characteristics, simple fabrication process and good compatibility with current complementary metal oxide semiconductor process (4). Many transition metal oxides have been reported to exhibit resistance switching phenomena, but until now, none of them has totally fulfilled the requirements of practical memory application and developing new functional materials is still meaningful.

In this paper, the resistive switching characteristics of MnO_x films were investigated and it showed promising resistive switching performances. The manganese oxide film was deposited by electron beam evaporation and sandwiched between two metal electrodes in the forms of Pt/MnO_x/Pt and Pt/MnO_x/Al. These devices exhibited stable resistive switching behavior under dc sweeping voltages as well as under pulse voltages. It was noticed that Pt/MnO_x/Al device had a better endurance performance than Pt/MnO_x/Pt, which was believed to be due to the aluminum electrode depriving oxygen from MnO_x thin film. Additionally, the Pt/MnO_x/Al device has good retention ability and its high/low resistance ratio could be over three orders magnitude. All the results suggested that Pt/MnO_x/Al had a potentiality for practical memory application. And the study of MnO_x could help to achieve a better understanding of resistive switching characteristics of TMOs.

Ex. 1016, Page 929

Experimentals

The MnO_x-based memory devices were fabricated as follows: (100)-oriented silicon wafers were chemically cleaned, and then a 150-nm-thick SiO_2 film was thermally grown. Ti layer and Pt film (with thickness of 10 and 50 nm respectively) were deposited sequentially onto the substrates by electron beam evaporation. Ti was used as an adhesive layer and Pt served as the bottom electrodes. After that, a 150-nm-thick MnO_x thin film was e-beam evaporated onto the Pt film using MnO_2 target in an ambient pressure of 2.6×10^{-6} Torr at room temperature (RT). Finally 50-nm-thick square-shaped Al and Pt top electrodes with dimensions arranging from 50×50 μm^2 to 800×800 μm^2 were deposited and defined by photolithography and lift-off process. X-ray photoelectron spectroscopy (XPS) analysis was utilized to investigate the chemical states of the MnO_x matrix. The current-voltage characteristics of the fabricated devices were measured by Keithley 4200-SCS semiconductor characterization system. The pulse test was performed with Agilent 81110A Pulse Pattern Generator and Tektronix DPO7104 digital phosphor oscilloscope.

Results and discussions

Both devices showed reversible resistive switching under sweeping voltages. Figure 1 shows the typical I-V curves. The inset is the schematic configuration of the measurement. The sweeping voltages were applied to top electrodes with bottom Pt electrode grounded. The pristine devices were in high resistance state (HRS). When the voltages swept from

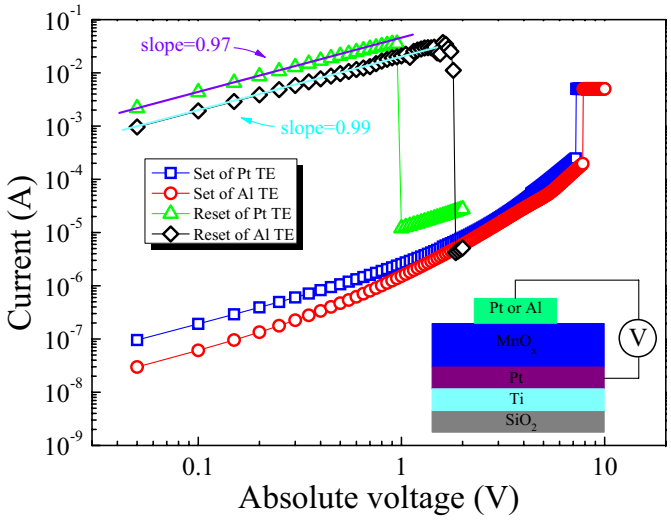


Figure 1. Double logarithmic plots of I-V curves of $Pt/MnO_x/Pt$ and $Pt/MnO_x/Al$ devices, the inset was the schematic plot of electrical measurement configuration.

zero to large, a drastic current jump occurred at about 8 V and the devices turned to low resistance state (LRS). During this process, the current was limited to a maximum value by the test equipment, in order to prevent the device from permanent breakdown. After the device turned to LRS, a sweeping voltage without current compliance was applied, and the current suddenly dropped, indicating the device turned to HRS. The switching between LRS and HRS could be done under arbitrary voltage polarity. Resistance switching under voltage pulse was also feasible. Set operation could be induced by a 100 ns pulse of 12 V and reset by a 200 ns pulse of 2 V.

In figure 1, the I-V curves of LRS has a slope very close to one, indicating that $\log(I) \propto \log(V)$, so the devices of LRS exhibits Ohmic conduction. Based on that, we proposed that resistive switching was caused by the formation and rupture of conducting filaments. According to the filamentary mechanism (5), under sufficient electrical field, defects of metallic ions or oxygen vacancies in the oxide matrix aligned to form highly conductive filaments. Once the conductive filaments formed, current mainly flows through these filaments and the devices acted as ordinary small value resistors. Current flowing through the filaments will generate Joule heating. When the current was large enough, local area would be heated to a very high temperature and the conducting filaments would break off as a result, and the device changed back to HRS.

Most of the Pt/MnO_x/Pt devices got breakdown after a few set/reset cycles, and then acted as stable low value resistors. While lots of Pt/MnO_x/Al devices still could function after more than 100 dc switching cycles when the devices were set by negative sweeping voltages. It was proposed more oxygen vacancies caused by aluminum depriving oxygen from MnO_x layer were the reason for the endurance difference. Aluminum was known to be able to get oxygen from TMOs to form oxides (6). According to the filamentary mechanism, during set process, conductive filaments propagated from the cathode and filaments near the anode were weaker (or less conductive) than the ones near the cathode. The filaments ruptured in a localized region near the anode during reset process, and recovered in the subsequent set process (5, 7). The rest of the filaments were preserved

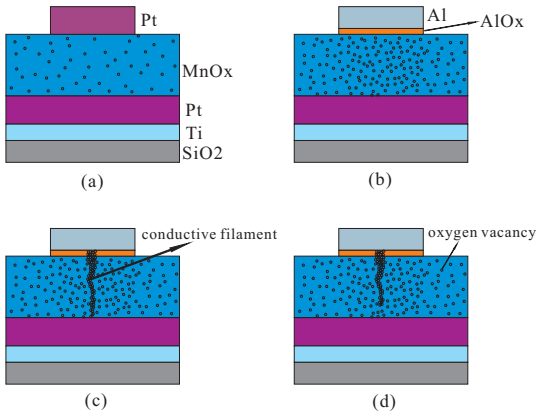


Figure 2. (a)The Pt/MnO_x/Pt device; (b) Aluminum could deprive oxyen from MnO_x and

made it have more oxygen vacancies; (c) When Al TE was applied with negative voltage, conductive filament grew from TE to BE; (d) conductive filament broke in the local area.

near anode during the whole switching process. As to our situation, the filaments broke and recovered in the MnO_x layers. This layer had been deprived oxygen by aluminum electrode and owned more oxygen vacancies, making conductive filaments' formation or recovery easier, so Pt/MnO_x/Al devices exhibited an improved endurance property.

The dependence of electrical characteristic on device sizes provides an insight to reveal the nature of the device. Figure 3(a) depicted the dependences of high and low resistance of Pt/MnO_x/Al devices on the size of top electrodes. With the increase of the size of top electrodes, the resistance values of HRS of the Pt/MnO_x/Al device got smaller, while the values of LRS showed no dependence on the device sizes, which was in agreement with the localized conductive filament mechanism. As shown in figure 3(b), the V_{set} (defined as the threshold voltage where set switching occurred) had little reliance on the cell area, while I_{set} (defined as the current when set occurred) increased with larger cell size. This indicated that the electrical field required to form conductive filaments was constant and the conduction of HRS was almost homogenous.

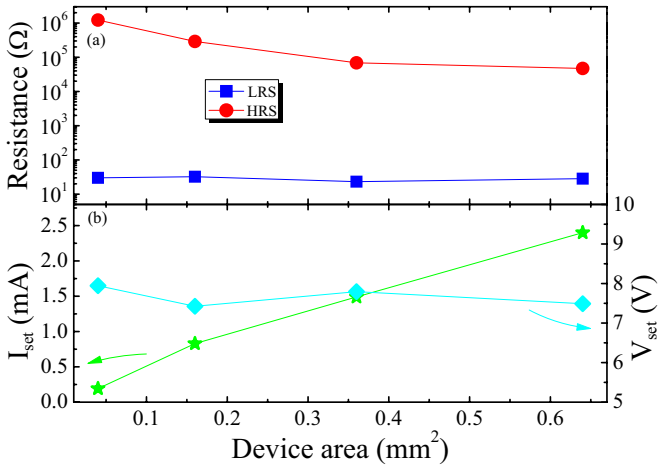


Figure 3. (a)The relationship of high and low resistances of Pt/MnO_x/Al devices with the device sizes; (b) The dependences of set current and set threshold voltage of Pt/MnO_x/Al on the device size.

Shown in figure 4 was the dependence of the resistance on the repetitive switching cycles of a 200×200 μm^2 Pt/MnO_x/Al device. The device was set by negative voltage and reset by positive voltage. Resistance values were obtained at 0.3 V. As can be seen from the experiment data, the low resistance values were distributed in a narrowly range around 100 Ω, while the high resistance values were around 1 MΩ. The narrow dispersion of resistance values guaranteed a sensing margin larger than three orders. The

inset of figure 4 depicted the distribution of the threshold voltages for set and reset (V_{set} and V_{reset}). There was no overlap between V_{set} and V_{reset}, which insured the Pt/MnO_x/Al device could be well operated in practical application.

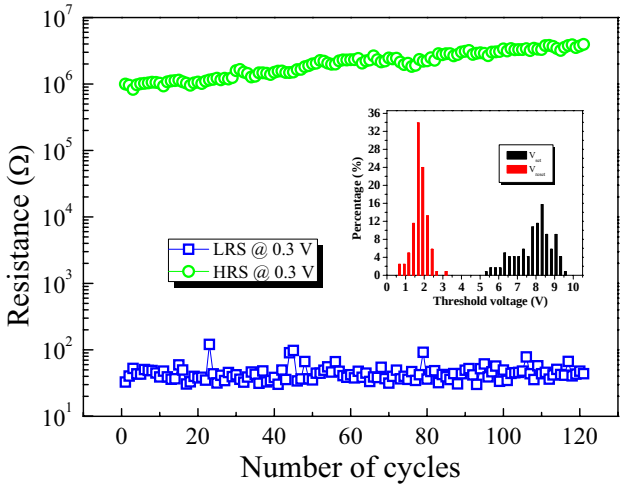


Figure 4. Resistance values versus cycle numbers for a 200×200 μm^2 Pt/MnO_x/Al device. The inset was the distribution of threshold voltages.

Figure 5 showed the time-dependent resistance evolution. The test at RT was done for over 10^5 s and retention at 85 ℃ was over 10^4 s. In both situations, the resistances of LRS exhibited little change, while those of HRS became larger, resulting in larger HRS/LRS ratios, enhanced the memory's reliability. The large difference of two HRS was caused by temperature.

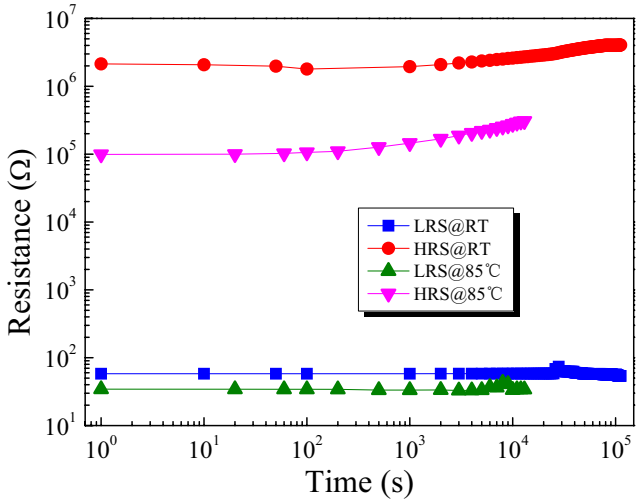


Figure 5 Retention performance of the Pt/MnO_x/Al device

Conclusion

The MnO_x thin film showed resistive switching phenomena. When Al was used as top electrode, the MnO_x-based device exhibited improved endurance performance. All the experimental results suggested Pt/MnO_x/Al had a potentiality for memory application.

References

1. S. Q. Liu, N. J. Wu, and A. Ignatiev, *Appl. Phys. Lett.*, **76,** 2749 (2000)
2. Beck, J. G. Bednorz, C. Gerber, C. Rossel and D. Widmer, *Appl. Phys. Lett.*, **77,** 139 (2000)
3. J. H. Jung, J. H. Kim, and T. W. Kim, M. S. Song, Y. H .Kim and S. Jin, *Appl. Phys. Lett.*, **89,** 122110 (2006)
4. I. G. Baek, M. S. Lee, S. Seo, M. J. Lee, D. H. Seo, D. S. Suh, J. C. Park, S. O. Park, H. S. Kim, I. K. Yoo, U-In Chung and J. T. Moon, *IEDM Tech. Dig. 2004*, p587 (2004)
5. K. M. Kim, B. J. Choi, Y. C. Shin, S. Choi and C. S. Hwang, *Appl. Phys. Lett.* **91**, 012907 (2007)
6. X. Wu, P. Zhou, J. Li, L. Y. Chen, H. B. Lu, Y. Y. Lin and T. A. Tang, *Appl. Phys. Lett.*, **90**, 183507 (2007)
7. K. M. Kim, B. J. Choi, B. W. Koo, S. Choi, D. S. Jeong and C. S. Hwang, *Electrochem. Solid-State Lett.*, **9**, G343 (2006)

ECS Transactions, 18 (1) 907-912 (2009)
10.1149/1.3096554 © The Electrochemical Society

FEA Based Design Optimization of Exciting Sensitivity for Micromachined Beam Ultrasonic Transducer

Yanan Han[1, 2], Chenguang Cai[2], Yang Liu[2], Antti Salo[2], Leon Xu[2], Jilong Xie[1]
[1] School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong Univeresity, Beijing 100044, China
[2] Nokia Research Center Beijing
No.5 Dong Huan Zhong Lu, BDA, Beijing, China 100176

Abstract: This paper presented a design optimization of piezoelectric micromachined ultrasonic transducer for its exciting sensitivity. A micromachined transducer is modeled as a large width-length ratio clamped-clamped beam, with the piezoelectric film to excite the beam vibration. The exciting sensitivity is taken as the optimization objective. Specifically, when the exciting voltage is fixed, the transducer structure is optimized to maximize the vibrating magnitude of the beam. The shape and the size of the piezoelectric layer are selected as the design parameters to optimize the exciting sensitivity of the transducer. In this paper, three different design variations of the piezoelectric layer structures are explored for the optimization. To perform the analysis and the subsequent optimization, the finite element models of the design options are created using ANSYS. The optimal design and the sensitivity analysis are performed; and the corresponding structural parameters of the piezoelectric film layer for optimal design are identified. Finally, we compare the optimization results of the exciting sensitivity for the large width-length ratio clamped-clamped beam based transducers of these three different piezoelectric layer structures; and the design method to improve the exciting sensitivity of the transducer is provided.

Keywords: micromachined beam ultrasonic transducer, exciting sensitivity, design optimazition, FEA

1. INTRODUCTION

Piezoactuation is one of the most common mechanisms in micro electromechanical systems (MEMS) devices. In conventional, a piezoactive film is deposited on a substrate of certain geometry, commonly a cantilever beam or a membrane. By applying an electrical field to the film, the resulting strain due to the direct piezoelectric effect causes the structure to bend, thus, performing useful work [1-2]. An ultrasonic piezoelectric transducer is just a typical application which can be used to generate ultrasound using the the piezoelectric film to excite the vibration of the transducer structure.

Normally, the piezoelectric film is designed to cover the whole top surface of the membrane or cantilever that is usually a Si-based film [3]. A typical micromachined piezoelectric ultrasonic transducer can be illustrated as in Figure 1 which was proposed by Z. Wang et al [4]. However the relatively low sensitivity of the structure limited its application fields such as microspeaker [5]. That is because the direction of the stress on center and boundary is opposite to each other from the stress distribution of the top surface of the membrane when the top surface of the membrane is totally covered with piezoelectric film and the exciting sensitivity will be negatively impacted [3]. Similar situation occurred in clamped-clamped beam when PZT film covers the whole top surface of Si layer, which is to be proved in details in next section.

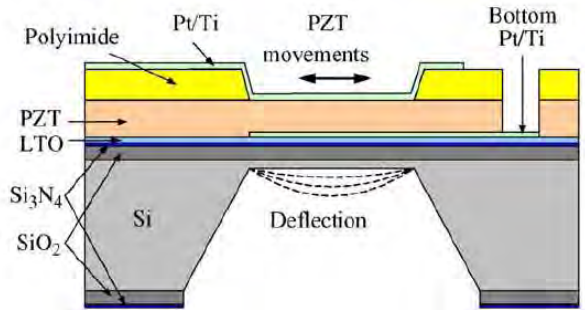


Figure 1 A typical micromachined piezoelectric ultrasonic transducer

2. THEORETICAL ANALYSIS AND FINITE ELEMENT ANALYSIS

To improve the exciting efficiency of the piezoelectric micromachined ultrasonic transducer, the bending behavior of the Si layer must be understood. For a conventional clamped-clamped beam under the uniform static pressure, the delfection and curvature variation along the length direction is shown in Figure 2, thereof $w_{\max}$ is the maxium deflection, $k_{\max}$ is the maxium curvature and l is the length of the beam. The curvature is negative in the clamped edge region but positive in the center region which means the opposite stress direction along the length direction similar to the membrane mentioned above.

Ex. 1016, Page 935

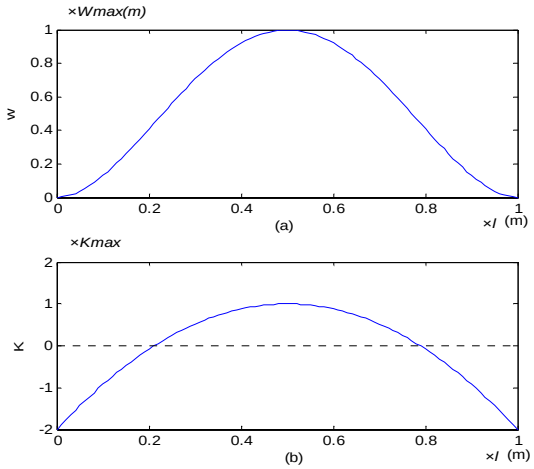


Figure 2 (a) The deflection variation along the length direction, (b) the curvature variation along the length direction

Finite element analysis is also done to verify above conclusion. In this paper, a clamped-clamped beam with the large width-length ratio 1 is modeled using ANSYS. The finite element model is shown in Figure 3. The width is 1.6mm that is the same as the length, while the thickness of Si layer is 15 μm and the thickness of PZT film is 1 μm. The exciting voltage of 10V is applied on the top electrode, while 0V on the bottom one. Note that, to show the FEA model used in this study, the thickness of the finite element model in Figure 3 is exaggerated.

Static analysis is then done; and Figure 4(a) and 4(b) illustrate the displacement in polar direction and stress distribution on top surface in the middle of the structure along its length direction relatively. It can be seen clearly that the compression is occurred in the clamped edge region but the tension is occurred in the center region in the middle of the structure.

Both the theoretical analysis and the finite element analysis give the conclusion that for the clamed-clamped beam with width-length ratio 1, opposite stress distribution occurs along the length direction which impacts the exciting sensitivity when PZT film covers the whole top surface of Si layer.

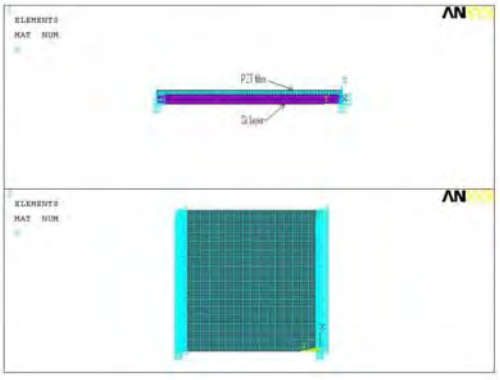


Figure 3 The finite element model of the clamped-clamped beam in ANSYS

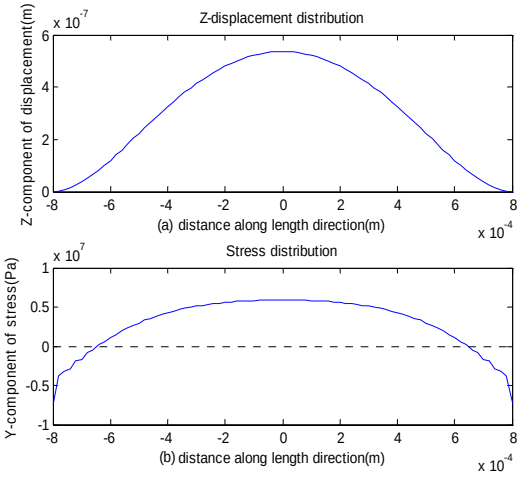


Figure 4 (a) The displacement in polar direction on top surface along its length direction, (b) the stress distribution on top surface along its length direction

Therefore, to achieve the maximum exciting sensitivity, the discrepancy of the stress between the center and the boundary on the Si layer should be maximized [3]. Out-of-phased driving in each PZT actuator or partially driving one of the PZT parts is required [6]. The paper pays attention to the latter by designing PZT film structures.

3. TRANSDUCER DESIGN

Based on the different direction on center and boundary of stress distribution, the shape and size of the piezoelectric layer is selected as the design parameters to optimize the exciting sensitivity of the transducer. In this paper, a clamped-clamped beam with the large width-length ratio 1 is modeled. Totally three different design variations of the piezoelectric film structures are explored for the optimization.

The first design option is to place the piezoelectric film layer on the middle of the clamped-clamped beam top surface as shown in Figure 5 (denoted as Structure 1). Hence, the piezoelectric film only produces the exciting stresses on the center of the beam.

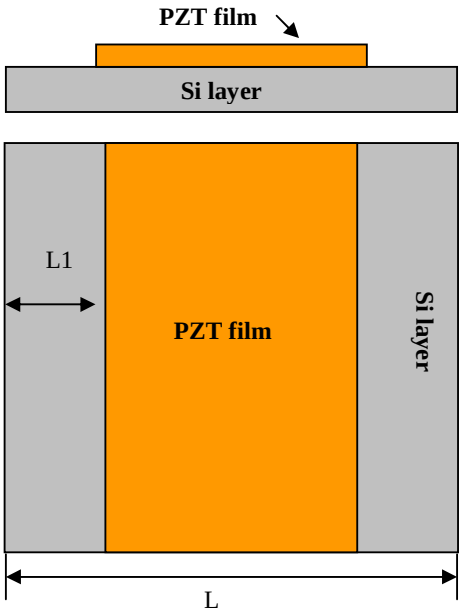
The second one is to place the piezoelectric film layer on two edges of the clamped-clamped beam, with the middle region of the piezoelectric film layer cut, illustrated in Figure 6 (denoted as Structure 2). In this case, the piezoelectric film only produces the exciting stresses on the edges of the beam.

The third one is the combination of the first two design options such that the exciting piezoelectric films are placed on both the boundaries and the middle of the clamped-clamped beam as shown in Figure 7 (denoted as Structure 3). With the opposite exciting voltages applied on the edges and the center of the beam, the piezoelectric film will separately produce opposite stresses.

For all the structures, ‘L’ means the length of the beam that is the same as the width. While ‘A’, ‘A1’ and

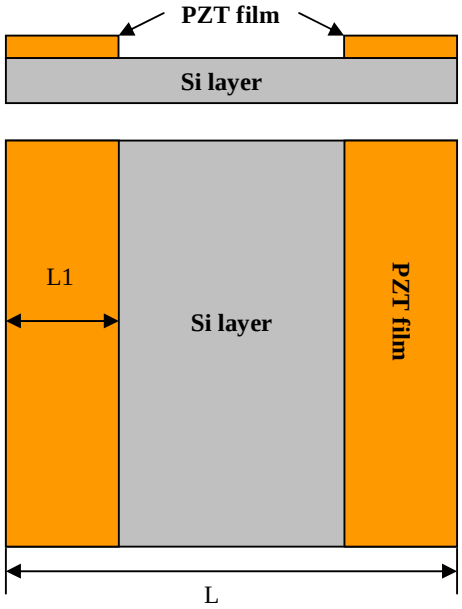
'A2' mean specific size ratio representing different structure configurations.

Using the finite element analysis, we try to identify the optimal design option among these 3 configurations and the corresponding optimal size for a specific structure which is to be presented in the following section.



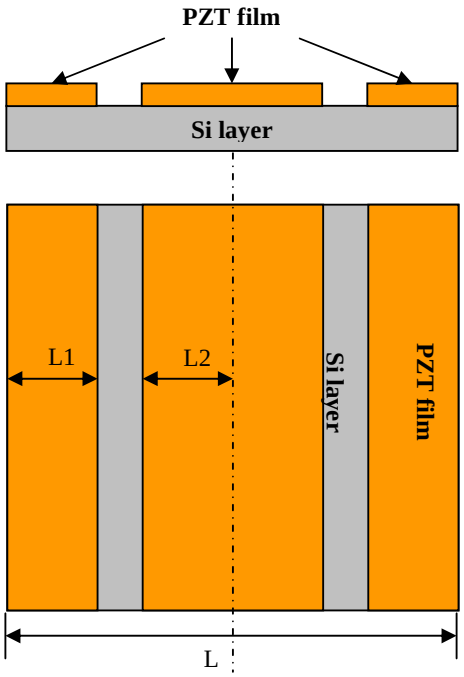
A=L1/L

Figure 5 The piezoelectric film located on the center of the beam



A=L1/L

Figure 6 The piezoelectric film located on the edges of the beam.



A1=L1/L, A2=L2/L

Figure 7 The piezoelectric films on both the edges and center of the beam

4. OPTIMAL RESULTS

To obtain the optimal structure and the optimal PZT film size, design optimization module of ANSYS software is used.

We neglected the damping effect in our analysis as a first approximation. Hence only the static responses were analyzed. Assuming that the transducers with the different piezoelectric film structures have the same damping, the gain of vibration magnitude of these different structures at their natural frequencies to the static displacements is the same. With this assumption, the maximal static displacement of the Si layer in polar direction is used as the optimization objective.

For all structures, the thickness of Si layer is chosen as 15μm; both of the width and length of the beam are chosen as 1.6mm. The exciting voltage of 10V is applied on the top electrode, while 0V on the bottom. But different thicknesses of piezoelectric film are chosen for optimization. The thicknesses are 3μm, 5μm, 7μm and 9μm respectively.

As to Structure 1 and Structure 2, the piezoelectric size effect on the maxium static displacement in polar direction is studied as shown in Figure 8 and Figure 9. 3μm and 5μm are chosen as the thicknesses of PZT film (denoted as H_PZT) for the sensitivity analysis. As the figures indicate, the same effect is found that the maxium static displacement in polar direction increases firstly and then reduces as the value of variable increases. So we can conclude that the maxium static displacement in polar direction exists definitely as long as A is chosen suitablely.

Ex. 1016, Page 937

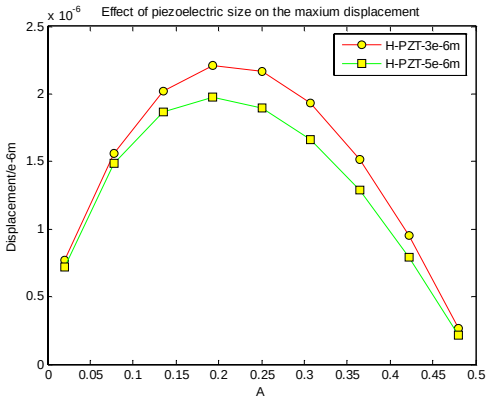


Figure 8 Displacement ***vs.*** A (structure 1)

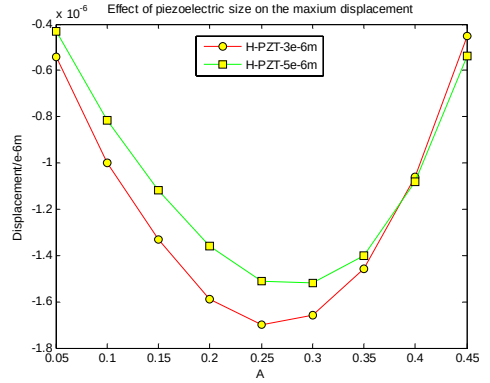


Figure 9 Displacement ***vs.*** A (structure 2)

In optimization, the objective function is the maxium static displacement in polar direction of the Si layer; the design variable is the PZT size denoted indirectly as A; taking into account practical application, the state variable is the natural frequency of the structure ranging from 30 kHz to 70 kHz. The sub-problem optimal method is chosen for the optimal method and the iteration time is set as 15.

Table 1 to Table 3 indicates the optimal results, showing maxium static displacement in polar direction with different piezoelectric thickness for the three stuctures and the corresponding A, A1 and A2. A=0 for Structure 1, A=0.5 for Structure 2 and A1+A2=0.5 for Structure 3 mean that the piezoelectric film covers the whole top surface of Si layer and the corresponding column is the maxium static displacement in polar direction at this condition. Optimal results show that the maxium displacement occurs while the thickness of PZT film is 3 μm for all structures. Specifically, for Structure 1 the maxium displacement is 2.22 μm with optimal A 0.20000; for Structure 2 the maxium displacement is 1.74 μm with optimal A 0.25000; for Structure 3 the maxium displacement is 3.18 μm with optimal A1 0.22450 and A2 0.26372. Moreover, the maxium displacement of these three structure configurations increases a lot with minimum 288.4% and maxium 609.8%, comparing with the maxium displacement when the PZT film covers the whole top surface of Si layer.

It turns out that Structure 3 is the best option, followed by Structure 1 and Structure 2. The structure that the PZT film covers the whole top surface of Si layer has the worst exciting sensitivity.

Figure 10 and Figure 11 illustrate the finite element results for the optimal Structure 3 with PZT thickness 3 μm. Figure10 is the displacement distribution field with the maxium displacement occurs in the center of the Si layer. Figure 11 shows the stress distribution field along length direction with minimum stress -30.7 MPa on the clamped edge and maxium 28.4 MPa in the center. The stress varies rapidly around the PZT film cutted region; and for the stress region with the same sign, the stress varies little.

Table1 Optimal results for structure 1

H_PZT(μm)	Optimal A	Maxium Displacement(μm)	A=0	Improvement (%)
3	0.20000	2.22	0.448	395.5
5	0.19100	1.98	0.375	428.0
7	0.16824	1.74	0.319	445.5
9	0.15540	1.51	0.277	445.1

Table2 Optimal results for structure 2

H_PZT(μm)	Optimal A	Maxium Displacement(μm)	A=0.5	Improvement (%)
3	0.25000	-1.74	0.448	288.4
5	0.27558	-1.55	0.375	313.3
7	0.29675	-1.37	0.319	329.5
9	0.31406	-1.20	0.277	333.2

Table3 Optimal results for structure 3

H_PZT(μm)	Optimal A1	Optimal A2	Maxium Displacement(μm)	A1+A2=0.5	Improvement (%)
3	0.22450	0.26372	3.18	0.448	609.8
5	0.21625	0.26537	2.51	0.375	569.3
7	0.22044	0.27218	1.99	0.319	523.8
9	0.20445	0.27194	1.59	0.277	474.0

Ex. 1016, Page 938

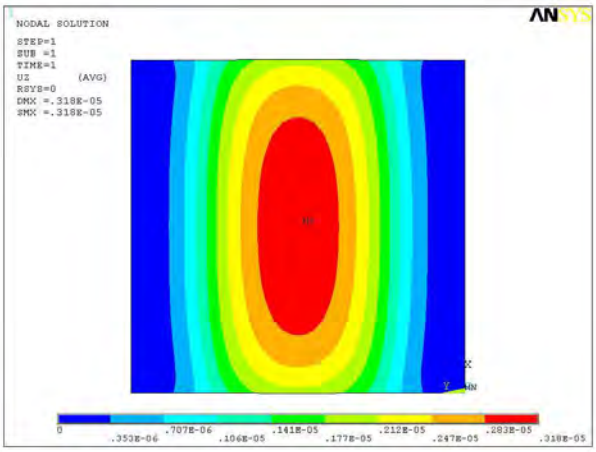


Figure 10 The optimal displacement distribution field

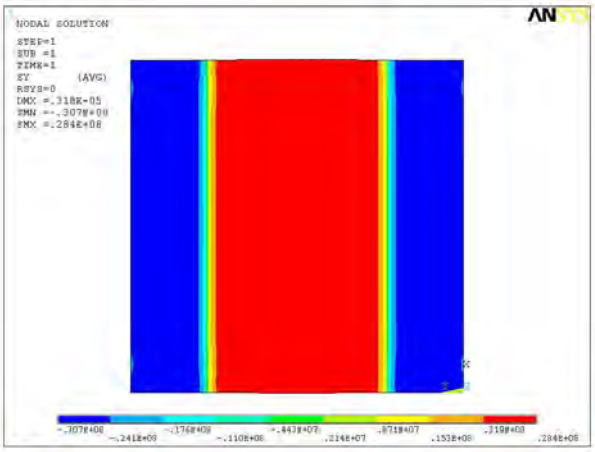


Figure 11 The optimal stress distribution field along length direction

5. CONCLUSION AND DISCUSSION

Based on the therotical analysis and finite element analysis, three different design variations of the piezoelectric film structures are explored for the optimization. ANSYS is ultilized; and the optimal PZT size coresponding to the different structures is obtained. Results show that Structure 3 (where the PZT film covered both the boundaries and the middle of the clamped-clamped beam) is the best option while the conventional structure that the PZT film covers the whole top surface of Si layer has the worst exciting sensitivity. The optimal maxium static displacement in polar direction shows a remarkable increase of 609.8% than the conventional clamep-clamped beam structure. In summary, placeing PZT film on both the boundaries and the middle of the clamped-clamped beam is the best structure to obtain optimal exciting sensitivity under the scope of this study.

The above results can be explained as follows: As we discussed above, to achieve the maximum exciting sensitivity, the discrepancy of the stress between the center and the boundary on the Si layer should be maximized. When the PZT film covers the entire top surface of Si layer, it creates a uniform offset in stress distribution on the entire surface. On the other hand, when the PZT film covers the part of the top surface of the Sil layer, it creates a stress offset only in the region where the PZT film locates, which may actually make the stress discrepancy even greater between the center and the boundary. Hence, it is understandable that the conventional structure (where the PZT film covers the entire top surface of Si layer) could create the smallest displacement compared to Structure 1, 2 and 3, as long as the size of the PZT film is appropriate. Note that when the size of PZT film is too small, our study indicates that the stress caused by the PZT excitation may not be big enough to create the needed stress discrepancy for vibration. Therefore, the optimal design for the best excitation sensitivity lies at a design point where the location and size of the PZT film should be optimized so that the stress discrepancy should be maximized between the center and the bouldary of the structure for the transducer discussed in this study.

Whereas Structure 1 and 2 create the better exciting senstivtity than the conventional structure, Structure 3 combines the advantages from both Structure 1 and 2. For Structure 3, with the opposite exciting voltages applied on the boundaries and the center of the beam, the piezoelectric film will separately produce opposite stresses. Therefore the discrepancy between the center and the boundary on the Si layer is enlarged. As a result, the clamped-clamped beam with PZT film covering both the boundaries and the middle of Si layer has the largest displacement.

When it comes to the specific size and position of the PZT film denoted as A1 and A2, the crucial point where the stress sign changes determines it indirectely. Based on the finite element analysis, when the PZT film covers the whole top surface of the Si layer, the crucial point is located in 0.368 mm and 1.232 mm of the Si layer. They are inverted as 0.23 and 0.77 as a propotion to the Si length 1.6 mm. From Table 3, the optimal A1 and A2 for PZT layer thickness 3 μm are 0.22450 and 0.26372 respectively. It comes to the conlusion that the optimal results make sense by removing the PZT part around the crucial point to eleminte the mutual liminations between the boundary and middle of the PZT layer. PZT layer thickness 5 μm, 7 μm and 9 μm have similar results. The cutted apart region is just around the crucial point.

Caution has to be made to further extend the conclusion from this study, which is our first step to optimize this MEMS-based transducer. Because we neglected the damping in the system, more sensitivity test regarding to the damping effect needs to be performed. Also, the harmonic analysis needs to be conducted in the future since the properties of the piezoelectric micromachined ultrasonic transducer as an ultrasound transmitter can be evaluated by output sound pressure level (SPL) at resonant frequency. Furthermore, pratical experiments should be conducted to verify analysis. However, as the first approximation, this study indeed

points out the direction of the transducer design for future design improvement for the best exciting sensitivity.

6. REFERNCES

[1] F. Akasheh, T. Myers, J. D. Fraser, S. Bose, A. Bandyopadhyay. Development of piezoelectric micromachined ultrasonic transducers, Sensors and Actuators A, 111 (2004) 275-287.

[2] H. Choi, A. Dalakoti, S. Bose, A. Bandyopadhyay. Influence of top electrode ddesign on pMUTs performance, Sensors and Actuators A, 135 (2007) 613-619.

[3] Chenguang Cai, Quayle Chen, Dujuan Zeng. Finite element analysis based design optimization of exciting efficiency for micromachined piezoelectric transducer, MicroNano2009-70265

[4] Z. Wang, W. Zhu, J. Miao, H. Zhu, C. Chao, O.K. Tan. Micromachined thick film piezoelectric ultrasonic transducer array, Sensors and Actuators A, 130-131 (2006) 485-490.

[5] S.C. Ko, Y.C. Kim, S.S. Lee, S.H. Choi, S.R. Kim, Micromachined piezoelectric membrane acoustic device, Sensors and Actuators A , 103 (2004)130–134.

[6] H. Lee, D. Kang and W.Moon. Design and fabrication of the high directional ultrasonic ranging sensor to enhance the spatial resolution, Transducers and Eurosensors ’07, the 14th International Conference on Solid-State Sensors, Actuators and Microsystems, 1303-1306.

[7] Linquan, Yao, Li Lu, Zhihong Wang, Weiguang Zhu, Member, IEEE, and Ying Dai. Exact solution of mutilayered piezoelectric diaphragms, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 50(10) (2003)1262-1271.

[8] P. Muralt, N. Ledermann, S. Baborowski, A. Barzegar, S. Gentil, B. Belgacem, S. Petitgrand, A. Bosseboeuf, and N. Setter. Piezoelectric Micromachined Ultrasonic Transducers Based on PZT Thin Films, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 52(12) (2005)2276-2288.

CHAPTER 9

SILICON TECHNOLOGY FOR SEMICONDUCTOR, ELECTRONIC, AND SOLAR APPLICATIONS

ECS Transactions, 18 (1) 915-924 (2009)
10.1149/1.3096555 © The Electrochemical Society

Polysilicon Production: Reaction Engineering and Scaleup Issues

P. A. Ramachandran
Department of Energy, Environmental and Chemical Engineering,
Washington University, St. Louis, MO, 63130, USA

In order to meet the increasing demand for solar energy utilization, polysilicon plants based on existing technologies are being built at an unprecedented rate and newer processes with better economic and environmental potentials are being evaluated. Reaction engineering, optimum reactor design and model based scaleup will therefore play a major role in achieving the dual objectives of capacity expansion and the reduction in cost of production. This paper reviews the various processes for manufacture of polysilicon and discusses the current state of the art in the reactor design and scaleup issues associated with polysilicon production, mainly for the emerging technology of fluid bed reactors.

Introduction

Silicon based photovoltaic (PV) cells are expected to play a major role in the utilization of solar energy and the market for polysilicon is expected to grow rapidly in the next five years. The projected worldwide capacity in 2012 is predicted to be about 200,000 m.t./year and individual plant capacity is expected to be of the order of 10,000 m.t./year, if not larger (1). Correspondingly, the share of power generation from PV is expected to increase and is already of the order of 3.8 GW in Germany. The target cost of production is set around $20-25 per kg. The silicon wasted from the semiconductor industry provided the raw material for solar cells in the early part of this century but the projected increase in capacity requires new and more efficient technologies to be developed which are specific to solar grade silicon (SoG) production. Also it may be noted that the purity requirements for SoG are less stringent than the semiconductor grade. All these factors have led to capacity expansion and development of several new ideas. The focus of the first part of the paper is, therefore, to provide a review of the current technologies for polysilicon production and some newer evolving technologies. These include the traditional Siemens reactor, the fluid bed reactor and the newer potential technologies such as the spouted bed reactor and the microwave assisted fluid beds. The reactor design and scaleup of these systems need a multiscale modeling methodology which encompasses the fields of chemical kinetics, particle nucleation and growth, hydrodynamic models and reactor models. The second part of the paper reviews such modeling pertinent to polysilicon production in a fluidized bed reactor which is emerging as a strong alternative route to the traditional Siemens reactor. Some aspects of scale up of these reactors are also discussed. It is hoped that the information and the key references provided here will be useful to industrial practitioners to further advance this field and provide an alternate sustainable energy source for the future.

The paper is organized as follows: The next section provides an overview of the various processes. The kinetics and thermodynamics of reactions leading to SoG silicon production are then discussed. This is followed by a review of experimental studies on

the fluid bed reactor for silane pyrolysis. The modeling and scale up aspects of fluid bed reactor then follow and some concluding remarks are provided at the end.

Overview of Processes and Technologies

In this section a description of various processes to produce SoG-Si and the associated reactors are described to provide an overview of the available or potential technologies.

Precursors

The starting material in most of these processes is metallurgical grade silicon which has a purity of 99% and require further purification since the purity requirement for SOG is around 99.9999%. The first step is the chlorination to produce $SiCl_4$

$$Si\ (met) + 2Cl_2 \rightarrow SiCl_4 \qquad [1]$$

which can then be converted to other precursors. The usual precursors are trichlorsilane (TCS) and silane. TCS is obtained by hydrogenation of $SiCl_4$ as shown below.

$$SiCl_4 + H_2 \rightarrow SiHCl_3 + HCl \qquad [2]$$

Silane (SiH_4) is produced by an additional process of hydrogen reduction of TCS

$$SiHCl_3 + 3\ H_2 \rightarrow SiH_4 + 3\ HCl \qquad [3]$$

or by a redistribution reaction with a solid catalyst. Other precursors being considered are bromine based precursors such as tribromosilane and direct reduction of $SiCl_4$ as discussed in the later sections of the paper. Various processes are then used to convert the precursors to poly-crystalline silicon as discussed in the following subsections. The poly is converted to single crystal Si by various melt growth processes. Commonly the Czochralski process is used for electronic grade applications while for solar applications the Bridgman, the ribbon growth, the edge defined film growth (EFG) and other processes are suggested since the quality requirements for solar grade is not as stringent as the electronic grade. The processes to convert the precursors to SoG are now reviewed.

Siemens Reactor

The Siemens reactor has been the workhorse for the semiconductor grade silicon. The process uses a bell jar reactor containing pure silicon rods which are heated to about 1100 deg C by electrical heating through the rods. A mixture of TCS and hydrogen is introduced into the bell jar reactor and the gases react to form Si on the rods. An alternative to the use of TCS is to use silane as a precursor. The silane decomposition requires a lower temperature (600-700 deg C) thereby reducing the energy costs. Hence the use of silane is more suitable for production of SoG, although an additional step of converting TCS to silane as per reaction #3 above is needed. Usually a mixture of silane and hydrogen is used as the reactor feed and the role of hydrogen is to prevent the homogenous nucleation of silane to Si in the vapor phase which leads to `dust' formation in the reactor and to achieve a preferential deposition on the seed rods. For further processing, the grown seed rods are crushed and single crystal silicon is grown usually by

the Bridgman growth process which is essentially a casting process rather than by the Czochralski process used in semiconductor industry. The modeling of Siemens reactor are presented, for example, by Yang (2) and del Coso et al. (3) and such studies are important to improve the process efficiency and to determine the optimum operating parameters. For further details, these works should be consulted.

Fluid Bed Reactor

The fluid bed reactor consists of seed particles of granular silicon of the size of 100μm in diameter. The bed is fluidized by a silicon precursor gas (usually TCS or silane) along with hydrogen. The Si is deposited by reaction of the precursors and a preferential deposition on the seed particles is sought for. Thus the seed particles grow to larger size of the order of 900μm and these large particles segregate at the bottom of the bed and are continuously withdrawn. Hence the fluid bed process can be operated in a continuous manner in contrast to the Siemens process. Other benefits claimed for the fluid bed process are: (i) Scaleup to larger capacity leading to lower production costs, (ii) less energy conversion and (iii) the absence of the crushing step of rods to grow crystals. The disadvantage is mainly due to the parasitic reaction in the gas phase which leads to gas phase nucleation and the formation of silicon dust or fines. Other drawbacks are the agglomeration of the growing particle leading to defluidization or unstable operation of the reactor and deposition on the walls of the reactor necessitating periodic shutdowns and cleaning. The scale up of fluid beds to larger sizes is also a challenging problem due to the complex hydrodynamic pattern in the bed. Hence the modeling and analysis of fluid beds (discussed subsequently in this paper) are expected to play a central role in further improvement of the process.

The common precursor used is TCS and hydrogen and the reaction leading to Si formation is:

$$SiHCl_3 + H_2 \rightarrow Si + 3\,HCl \qquad [4]$$

The reaction is reversible and the maximum conversion occurs around 1100-1150K and this is the temperature range for the process. More recently the pyrolysis of silane

$$SiH_4 \rightarrow Si + 2\,H_2 \qquad [5]$$

appears to be gaining ground and the temperature required are compartively low, of the order of 600 deg C. The formation of fines is, however, much more here than in the TCS reaction since the reaction is practically irreversible under the operating temperature and pressure. Other precursor being considered is TBS, tribromosilane (1). The decomposition temperature for this reaction is around 800 deg C and the reaction does not lead to significant fine formation unlike TCS or silane. However, bromine needed to produce TBS is more expensive than chlorine and the recovery of HBr and recycle of bromine is expected to add additional costs.

Zinc Reduction of $SiCl_4$

Zinc reduction of silicon precursor such as $SiCl_4$ is being investigated as an alternative way of producing SoG. An earliest work was the JPL report study in reference (4). The process is now currently being investigated by Chisso Corp, Japan (5), the so called JSS

(Japan solar silicon) method and by other companies (6). The advantage of this process is the direct use of $SiCl_4$ as the raw material and no further reaction of $SiCl_4$ to produce other precursors is needed. The reaction chemistry can be represented as:

$$SiCl_4\ (g) + 2\ Zn\ (vap) \rightarrow Si\ (s) + 2\ ZnCl_2\ (g) \qquad [6]$$

The process is followed by recovery zinc from $ZnCl_2$ by either hydrogen reduction or electrolytic process. Thus overall process of going from met-Si to SoG-Si is envisaged as a closed loop system with no waste. The reduction reaction itself can be carried out in a fluid bed reactor. Detailed reaction engineering analysis of the process does not appear to have been published and will be useful for further development of this potential process. The models for fluid beds presented in a later section of this paper can be a useful starting point for such a modeling analysis.

Upgraded Metalurgical Silicon

Direct upgrading of metallurgical silicon (UMG) is being touted as a simple and direct route obviating the need to go through the many chemical steps listed above. The information is proprietary and the data for the evaluation of the technology for the process are scanty in the open literature. Basically, the steps involved are similar to the conventional purifying processes in steel industry, *viz*., slag treatment, leaching and solidification (1). Typical impurities are 1.5 ppmw of phosphorus and 0.4 ppmw of boron. Blending of UMG with SoG produced from other routes is also suggested to bring down the overall cost of PV modules.

Thermodynamics and Kinetics of Reactions

Thermodynamics and kinetics are the two of the five pillars of chemical reaction engineering as advocated by Fogler (7). Thermodynamic studies indicate the favorable range of operating conditions while the kinetics show how fast the reaction can proceed. The thermodynamic behavior of Si-H system has been investigated in detail by Meng and Hunt (8). The corresponding study for the Si-H-Cl system has been done by Hunt (9). This work also reports some data for Si-Br-H system which is useful for the evaluation of the tribromosilane process.

The kinetics can be modeled at three hierarchical levels. (A.) Simple single step or two step reaction schemes, (B)_ Microkinetic models and (C.) Detailed chemistry models.

Also the reaction scheme is classified into homogeneous and heterogeneous reactions. The homogeneous reactions take place in the gas phase and are mainly responsible for the fine or dust formation and hence are undesirable. The heterogeneous reactions take place on the seed or growing particles.

Homogeneous Reactions: The simple reaction model for homogeneous reaction of silane is usually represented as a first order reaction with the rate constant following an Arrhenius law. A two step scheme used in some modeling studies and here the reaction mechanism assumes the formation of silyene (SiH_2) as an intermediate:

$$SiH_4 \rightarrow SiH_2 + H_2 \qquad [7]$$

This reaction is then followed by reaction of silyene on the solid surface of the growing silicon particles:

$$SiH_2 \rightarrow Si + H_2 \qquad [8]$$

The kinetic parameters (frequency factors and the activation energies) are provided in the paper by Furasawa et al. (10) and are useful for reactor models using simple kinetics. The pressure dependency of the rate constant for silane pyrolysis has been studied based on a molecular model by Meyerson and Jasinski. (11). They used the Rice, Ramsperger, Kassel and Marcus (RRKM) theory to predict the rate constant over a wide range of pressures. The results are more qualitative than exact but nevertheless useful to predict the reactor performance for potential operation of the reactor at elevated pressures.

Microkinetic models (12) for homogenous reactions start with the first step which is the formation of the silyene (reaction [7] above) which is then followed by the formation of dimeric and polymeric silicon species such as Si_2H_6, Si_3H_8 etc. A typical reaction is indicated here.

$$SiH_4 + SiH_2 \rightarrow Si_2H_6 \qquad [9]$$

These species can further polymerize or decompose to form Si. One of the postulate is that these types of reactions leading to the formation of higher Si-polymers have a key role in the formation of the fines as discussed later in the paper.

Detailed kinetic models have been proposed by Coltrin et al. (13) and Girchick et al. (14) and these are useful to understand the relative rates of homogenous and heterogeneous reactions in a more fundamental manner. For example, Coltrin et al. considered a reaction scheme with 27 reactions and 15 species which included the various silicon hydrides of the formula Si_nH_m upto n=3 along with Si_2 and Si_3. A more detailed kinetic study was performed by Girchick who included polymeric species upto the order of n=10. Since the mechanism of homogeneous nucleation and fine formation depends on these species, the detailed kinetic models are apparently very important. However, for applications to reactor design, some model reduction methods are needed since it is computationally prohibitive to couple complex hydrodynamics with a detailed chemistry. Although a number of tools are available in chemical engineering literature, these techniques do not appear to have been used to derive reduced order kinetic models for this process.

Heterogenoues Reactions: For heterogeneous reactions, the simple models are extensions of the Langmuir-Hinsehelwood models widely used in heterogeneous catalysis and reaction engineering. Here the mechanism assumed is the adsorption of the reacting species on the particle surface followed by its decomposition according to:

$$SiH_4\,(g) \rightarrow SiH_4\,(ads) \rightarrow Si\ (s) + H_2\,(g) \qquad [10]$$

For a two step reaction scheme, the adsorption of SiH_2 followed by surface reaction is also included in the kinetic scheme. The kinetic model based on this mechanism has been proposed by Furasawa et al. (10) and is widely used in modeling studies and the rate

constants and temperature dependency of rate have been reported in their work. For heterogeneous reactions, the detailed microkinetic models do not appear to have been studied.

Fine Formation Mechanism: The main mode for silane decomposition is by two pathways as discussed by Lai (15). One is the homogeneous pathway and the other is the heterogeneous pathway and Lai et al (16) indicated that the homogeneous pathway is responsible for the fine formation. In their model, SiH_4 decomposes in the gas phase to form Si vapor. It may be noted that the vapor pressure of Si is low but is included in the model. Once the Si vapor concentration exceeds the supersaturation value, the Si vapor nucleates into Si nuclei which then coagulates and coalesce into fine particles. The fines can also be captured by the seed particles (the scavenging mechanism). Lai et al. adopted this mechanism and coupled it with a phenomenological model for fluid beds.

In some recent studies this mechanism has been challenged and a contrarian view of fines forming from polymeric Si hydrides has been proposed (14). These polymeric species have the general formula, Si_nH_m. Detailed numerical simulation based on this model has been proposed by Girshick et al.(14). These authors combined a detailed kinetic scheme containing up to polymeric species of n=10 and an aerosol dynamic model to predict the fine concentration and the size of the fines. The method of moments was used in the aerosol model rather than a detailed population balance model. The model was for a batch reactor and shows qualitative agreement with the published experimental results. Since the fine formation is usually detrimental to the process economics, a more detailed study and experimental confirmation of the mechanism of nucleation and fine formation would be useful.

Fluid Bed Reactors: Experimental Studies

Thermal decomposition of silane in a fluid bed appears to be gaining ground in the production of SoG due to several advantages as discussed in the earlier section. In this section some key experimental studies on this system is presented together with a summary of the main findings. It may be noted that most of the published data are for small diameter reactor and the direct use of such data to design a larger diameter reactor is not possible and experimental studies in a pilot scale is usually needed. The modeling studies presented in the next section provide some guidelines but the successful scaleup is a complex problem and is more of an art rather than an exact science.

Experimental data for growth of large diameter particles (350μm) was obtained by Terjero-Ezpeleta et al. (17). The range of conditions studied were as follows: Bed diameter 5cm; bed height 30 cm, Temperature 923K, mole fraction of silane 1 to 10%. Carrier gas: nitrogen and hydrogen. The growth rate in the range of 25 μm /hr is typically observed in most of the studies. These growth rates agree with the values reported in an earlier study by Hsu et al. (18). It was observed that the particle size distribution did not change with time although the average diameter increases from 350 to 900 μm in about 42 hours. This indicates that the chemical vapor deposition occurs on particles of all the sizes in an equal manner. The fines are all elutriated out of the bed and the particle size distribution obtained experimentally was only representative of the bed. The size distribution of the dust elutriated from the column showed a mean size of 10 μm and a variance of 2μm.

Experimental studies were also reported in a paper by Caussat et al (19). The conditions of the studies were as follows: Bed diameter = 5.3 cm; Bed height = 63cm; carrier gas = nitrogen; seed diameter = 36 μm of alumina. In addition to studies with silane, disilane was also used as the reactant gas in order to ascertain the differences in rate and fine formation in the two systems. The main results are summarized below. The fluidization behavior when silane was introduced in the bed changed dramatically as revealed by the pressure fluctuations and the analysis of the pressure fluctuations with fast Fourier transform. This indicates that the particle surface characteristic may affect the quality of the fluidization. A large temperature variation in the bed of the order of 60 deg C was observed between the bottom and top. The fines were essentially formed in the walls and also in the freeboard regions in the bed. This is disagreement with the earlier studies of Hsu et al. (18), Lai et al. (16) who suggested that the fines are formed by homogeneous reaction of silane in the bubble phase of the fluid bed (also see earlier discussion on fine formation). Significant agglomeration of the seed was also observed when the inlet silane concentrations exceeded 20%.

Fluid Bed Reactors for SoG: Modeling Studies

Some general aspects of fluid bed modeling are first provided below to form a perspective for readers not familiar with the field. Fluid bed models can be classified into two types;

(i) Phenomenological models and (ii) Detailed computational models.

Phenomenological models assume that the bed can be divided into two regions (a) a bubble phase with a large gas voidage containing very little solids and (b) an emulsion phase which contains most of the seed particles. Hence the general terminology of `two phase models' are used to denote such models. The two phase phenomenological models can further be classified into of four types (20): (A). Patridge and Rowe model (B). Kunii-Levenspiel model (C) Modified Kunii-Levenspiel model and (D) Kato-Wen model. These models essentially differ in the manner by which the flow patterns in the bubble and emulsion phase are treated and whether the presence of a cloud phase between the bubble and emulsion phase is explicitly included or not. All the models assume that mass is exchanged between the bubble and the emulsion phase and the mass exchange depends on the bubble size.

Detailed hydrodynamic models use a point to point momentum balance to compute the local velocity and void fraction and do not explicitly assume the presence of a separate bubble phase. Detailed models can be of two types.

(i) Euler-Euler model which treats the two phases (gas and solid) as an interpenetrating continuum and solves for the momentum equations for both gas and solid phases. The model equations are essentially the gas and the solid phase continuity equations, momentum balances with phase interaction terms and the energy balances. These equations are supplemented by a large number of closures (the constitutive relations) for the phase interaction terms. It may be noted that the model predictions are sensitive to the parameters used for the closure.

(ii) Discrete element model (DEM) or discrete particle model (DPM) which treats the gas as a continuum but treats the solid phase by motion of individual particles modeled

Ex. 1016, Page 949

by the Newton's second law of motion. These type of models are also known as Euler-Lagrangian models. Although it is physically more realistic to treat the solids as a discrete phase rather than continuous, this model is computationally more difficult than the Euler-Euler model. Also no study of application of DEM to silane pyrolysis has been reported in the literature and a comparative study of the two approaches would be useful.

With this introductory background, we now present the review of various models presented in the literature for silane pyrolysis.

Phenomenoligal Models:

Caussat et al. (21) tested the four phenomenological models listed above to study silane pyrolysis in fluid beds. The Kato and Wen model showed closer agreement with the experimental data. Detailed homogenous reactions were then added to the Kato-Wen model. The major findings were as follows; Silane conversion was essentially complete in the first few cm of the bed. The fines are essentially formed by wall deposition of the various polymeric Si species leaving the bed rather than by homogeneous decomposition of silane followed by nucleation. This is in disagreement with other studies where the homogenous reaction of silane (mainly in the bubble phase) is considered to be responsible for the fine formation.

Pina et al. (22) developed a model for a spouted bed reactor based on the two phase model for fluid beds. **Spouted bed** is a modification of the fluid bed where the inlet gas is introduced by a nozzle somewhat above the bed. The particles are fluidized in the region above this nozzle while in the bottom part, the solids circulate as a spout. Modeling of such a reactor needs a spout-emulsion transport region for the bottom part and a bubble-emulsion region for the fluidized part and such a model developed by Pina et al. (22). Their model is similar to that proposed by Lai et al (16) earlier who consider a jet-emulsion and a bubble-emulsion region. The main difference is that a detailed population balance model was used to predict the particle size distribution in the bed for both the fines and the larger seed particles while Lai used the method of moments only for the fines. In either of the two approaches, the model for the particle size distribution requires two additional parameters: scavenging constant and the agglomeration constant as discussed earlier. These parameters were fitted to the experimental data in the work of Pina et al.. The comparisons were done for five experiments and these parameters had to be adjusted for each case. Hence the models for particle size prediction do not appear to be predictive and require further development. The fluid bed model itself predicts that the silane conversion is complete in the first 10% of the bed. The model predictions were sensitive to the exchange coefficients needed in the two phase model; In particular the sensitivity was greater for the spout-emulsion exchange coefficient compared to the bubble-emulsion exchange coefficient.

Huang et al. (23) modeled the behavior of a large diameter fluid bed reactor (80cm) again on the basis of the bubble-emulsion phenomenological model. The major modification was that the emulsion phase was considered to be a number of interconnected well mixed tanks rather than one single well mixed tank. The model was compared to experimental data for two runs obtained from an industrial pilot plant and reasonable agreement was found between the model predictions of silane conversion and selectivity to fines. In view of the space limitations, these comparisons and other details

are not presented here and will be published in a subsequent paper. Some of the important findings of this study are however summarized in the following paragraph..

Although one to one comparison was not done with the previous models on small diameter beds, it appears that the fine formation rate is larger in larger diameter fluid beds. This may have important bearing during scaleup. The reason attributed to this effect was the poorer bubble to emulsion mass transfer coefficient in larger diameter beds. Additional factors leading to higher fine formation are the low solid holdup in the bubbles, a larger volume for homogeneous nucleation in the bubbles and the lack of scavenging of fines in the bubble phase. The concentration of silane in the emulsion phase was significantly lower than that in the bubble phase indicating that the bubble-emulsion mass transfer resistance is a controlling resistance. Also the silane conversion was almost complete in the lower sections of the bed in the emulsion phase but complete conversion was not achieved in the bubble phase in contrast to the predictions in smaller diameter beds. Based on the model further parametric effects of various operating parameters were examined. The silane conversion increased slightly with the increase in the inlet silane mole fraction but the quantity of fines produced increased as well. Hence there is an optimum value for the inlet concentration of silane for which the productivity will be a maximum. The effect of wall temperature profiling on the productivity was also examined. These parametric studies are useful for reactor optimization and to minimize wall deposition. The model itself coupled with pilot scale studies provides a solid foundation for scaleup to larger sized plants.

Detailed Fluid Dynamic Models

Euler-Euler continuum approach has been investigated by O'Brien et al. (24) The MFIX code developed by NETL, Morgantown, WV was used for this simulation and the silane decomposition in a 5cm diameter and 63 cm tall reactor was studied. It was found that the reactions were essentially complete in the 4cm region above the distributor. The model predictions for macroscopic quantities (such as the H2 production rate) agreed with the data reported by Caussat et al (19) but this is not a very accurate test for a detailed computational model. More elaborate pointwise (microscopic level) data for comparison of the detailed model are needed rather than comparing the overall performance An interesting observation of this study was a 2-D axisymmetric model showed significant difference compared a full 3-D simulation. Due to computational costs, the phenomenological model would still be used for large scale reactor design but the detailed fluid models are expected to provide a guiding light for phenomenological models, especially for the mass exchange constants.

In a related study Xie et al (25) used the *in situ adaptive tabulation* (ISAT) and linked this to the MFIX model. ISAT is essentially a scheme to simplify the computations of a detailed kinetic model so that it can be used in detailed fluid dynamic codes such as MFIX. The work of Xie et al., however, used only a simple isothermal reaction scheme with only five kinetic steps and their goal was mainly to test the logistics of linking ISAT to a Euler-Euler two phase fluid bed model. Future studies should address the implementation of more detailed kinetic models with MFIX code in order to predict more accurately the fine formation etc.

Microwave Assisted Fluid Beds: The fines and wall depositions can be reduced if the particles are held at a higher temperature than the gas and the walls and one suggested

Ex. 1016, Page 951

strategy is the microwave assisted fluid bed, since silicon is highly microwave absorbing material Some information on using such a reactor with TCS as the precursor is available in the literature (26) although detailed models do not appear to have been proposed.

In **conclusion,** a number of promising technologies are now available for producing SoG including the fluid bed process with silane as the precursor. Reaction engineering analysis has also progressed significantly in the last ten years and a multiscale modeling approach seems to be emerging. For fluid bed reactors, the bubble diameter and the phase exchange coefficients appear to be the key parameters of importance in scaleup and a model based scaleup is expected to play a key role in the overall process development.

References

1. G. Parkinson, *Chem. Eng. Progress*, pp 8-11, August (2008)
2. Yang, Yu Bo, *D. Sc. Thesis, Washington University* (1988).
3. G. del Coso, I. Tobias, C. Canizo, A., Luque, *J. Crystal Growth*, v299, pp165-170 (2007).
4. J. M. Blocher, M. F. Browning, W. J. Wilson and D. C. Carmichael, *ERDA/JPL Report*, 954339=76/1 (1976).
5. Chisso Corporation Report, http://techon.nikkeibp.co.jp/english
6. S. Takayuli and I. Nobuo, *Annual Report of the Ceramics Research Laboratory*, Nagoya Institute of Technology, v 4, pp37-42 (2005)
7. S. Fogler, *Elements of Chemical Reaction Engineering*, Prentice Hall (2006)
8. T. Meng and L. P. Hunt, *J. Electrochem. Soc.*, v139, pp806-809, (1992)
9. L. P. Hunt, *J. Electrochem. Soc*., v135, pp206-209, (1988)
10. T. Furasawa, T. Kojima and H. Hiroha, *Chem. Eng. Sci*., v43, pp2037-2042, (1988).
11. B. S., Meyerson and J. M. Jasinski, *J. Appl. Phys*, v61, pp785-788, (1987).
12. C. J. Giunta, R. J. Mac Curdy, J. D. Chapple-Solol and R. G. Gordon, *J. Appl. Phys*, v62, pp1062-1067, (1990).
13. M. E. Coltrin, R. J. Kee and J. A. Miller,. *J. Electro. Chem. Soc*. v133, 3-8 (1986).
14. S. L. Girshick, M. T. Swihart, S. M. Suh, M.R. Mahajan and S.Nijhawan, *J. Electro. Chem. Soc*. v147, 2303-2311 (2000).
15. S. Lai, *D. Sc. Thesis, Washington University in St. Louis*, (1987)
16. S. Lai, M.P. Dudukovic, and P. A. Ramachandran, *Chem. Eng. Sci.*, v50, pp3615-3623, (1995)
17. M. P. Tejero-Ezpeleta, S. Buchholz, L. Mleczko, *Can. J. Chem. Eng*, v82,pp 520-529, (2004).
18. G. Hsu, R. Hogle, N. Rohtagi, and A. Morrison, *J. Electro. Chem. Soc*. v131, 2-8 (1984).
19. B. Caussat, M. Hemati and J. P. Couderc, *Chem. Eng. Sci.*, v50, pp3615-3623, (1995).
20. D. Kunii and O., Levenspiel, *Fluidization Engineering*, Butterworh-Heinamann (1991).
21. B. Caussat, M. Hemati and J. P. Couderc, *Chem. Eng. Sci.*, v50, pp3625-3635, (1995).
22. J. Pina, V. Bucala, N. S. Schbib, P. Ege and H. Ignacio de Lasa, *Int. J. Chem. Reaction Engg*.: Vol. 4: A9. (2006).
23. Y. Huang, P. A. Ramachandran and M. P. Dudukovic, AIChE, Annual Meeting, Salt Lake City , Paper #463b , (2006).
24. T. J. O'Brien, M. Syamlal, C. Geunther, *Third Int. Conference on CFD in the Minerals and Process Industries, CSIRO*, Melbourne, Australia, pp469-4764, (2003).
25. N. Xie, F. Battaglia, and R. D. Fox, *Combustion Theory and Modeling*, v8, pp195-209, (2004).
26. H. Y. Kim, Y. M. Song, T. Y. Jeon, D. H, Kwon, L. S. Lee, and D. S. Park, *US Patent*, 5,374,413, (1994).

ECS Transactions, 18 (1) 925-933 (2009)
10.1149/1.3096556 © The Electrochemical Society

Heat and Impurity Transfer Mechanisms of Czochralski and Directional Solidification Processes

K. Kakimoto, X. J. Chen, L. J. Liu, H. Miyazawa, H. Matsuo,
S. Nakano, S. Hisamatsu, Y. Kangawa

Research Institute for Applied Mechanics, Kyushu University, Japan

In this paper, results of global computation in a three-dimensional configuration for transverse magnetic field-applied Czochralski method to study distributions of temperature and impurity concentration are presented. The analysis includes a re-melting and segregation model at an interface between a crystal and the melt. Deflection of an interface between a crystal and the melt is also taken into account.

Time-dependent global analysis of the solidification process of a silicon ingot for photovoltaic is described in this paper. This model includes distributions of dislocation, impurity and point defects for a quasi-single silicon ingot grown by the solidification method. Heat and mass transfers in a square crucible are also discussed.

Introduction

Silicon crystals have been key materials in information and solar-energy technology. The CZ growth method has been the most widely used technique over the past several decades for growing high-quality bulk crystals of silicon. Directional solidification is a widely used method for producing multicrystalline silicon for solar cells. Crystalline Si produced by this method is the most widely used material in solar cells and is now used in about 60% of PV modules worldwide. Recent requirements of large-diameter wafers of high quality for the semiconductor industry together with the requirement of cost-effective wafers with a long carrier lifetime for the solar market have led to the development of crystal growth furnace and whole growth systems as well as to control of the growth process. The thermal field in and impurities incorporated into a grown crystal, the melt-crystal interface and the melt flow motion in the crucible have significant effects on the formation of micro-defects and crystal quality in a crystal growth process. It is therefore important to elucidate and control the heat and mass transfers and the distributions of impurities as well as the melt-crystal interface shape in a crystal growth process. With the increasing capacity of modern computers and progress of computation technology, application of numerical simulation has become an essential and effective tool for the design and optimization of a crystal growth process to improve crystal quality.

Over the past two decades, there have been many works on numerical analyses of bulk crystal growth processes using various models (1-10). Most of these models can be generally divided into two types: local models (1-4) and 2D global models (5-8). However, since all constituents of a growth furnace are closely related to each other through different forms of heat transfer by melt convection, solid conduction and thermal

Ex. 1016, Page 953

radiation, the growth furnace is a highly nonlinear and strongly conjugated thermal system. Quantitative analysis of heat transport throughout the entire furnace is thus required. On the other hand, the melt flow in a crucible and the thermal field within the growth furnace are principally three-dimensional, especially in cases under the influence of asymmetric external fields, such as in a transverse magnetic field-applied Czochralski method (TMCZ) or electromagnetic field-applied Czochralski method (EMCZ) or when using a square crucible in directional solidification. Therefore, 3D global analysis of a growth system is necessary for a quantitative understanding of realistic phenomena and the physics of the growth process and for process improvement of the growth process.

A coupled 2D/3D model (9, 10) was developed to solve these problems. In this kind of model, local 3D analysis is carried out with thermal boundary conditions obtained from a quasi-2D global modeling. Obviously, the 3D effects of the thermal system, especially the 3D melt flow, cannot be taken into account in the quasi-2D global modeling, and the thermal boundary conditions for the 3D local analysis can only be given in two dimensional. Furthermore, how to give the thermal boundary conditions for the 3D local analysis from the solution of quasi-2D global modeling remains a problem, since it has been reported that the type of thermal boundary conditions has a strong influence on the 3D results of interface shape and on the melt convection in a crucible (9). To solve the above-mentioned difficulties in the previous models, we developed a numerical simulator implemented with a novel 3D global model of heat transfer, which makes 3D global modeling feasible with moderate requirements of computer resources and computation time (11, 12). All convective and conductive heat transfers, radiative heat exchanges between diffuse surfaces and Navier-Stokes equations for the melt phase are coupled and solved together with a finite volume method in a 3D way. The developed simulator has been successfully used to quantitatively predict the 3D features of crystal growth.

The simulation code has also been successfully been used for a variety of crystal growth processes, such as solution growth for GaN (13, 14), PVT method for SiC (15, 16), and CZ growth and directional solidification processes for Si crystals (11, 12, 17-23). In this paper, application of the simulator to a CZ growth process and a directional solidification process for Si crystals, the two major production methods for crystalline Si for solar cells, is introduced. Some typical results are presented to show the importance and effectiveness of numerical simulation in understanding the physics, design and optimization of crystal growth processes.

Numerical model and governing equations of the crystal growth system

Typical configurations of a CZ furnace and a directional solidification (DS) furnace are illustrated in Figs. 1 and 2, respectively. For the DS furnace, a cylindrical or a square crucible was used in this study. All of the constituents of the furnaces are subdivided into a set of block regions as shown in the figures. Each block region is covered by a structured grid. When 3D global modeling is carried out with moderate requirements of computer memory and computation time, a mixed 2D/3D finite volume scheme is used (11). Following this scheme, the components in the core region of the furnaces, namely, the 3D domain, are discretized in a three-dimensional way, while the domains away from the core region, namely, the 2D domain, are discretized in a two-dimensional way. In Fig. 1, the 3D domain includes the crystal, melt, crucible and heater. The other regions in the furnace are included in the 2D domain. Precise calculation of view factors in the radiation modeling is an important part of the model with such a space discretization scheme.

Ex. 1016, Page 954

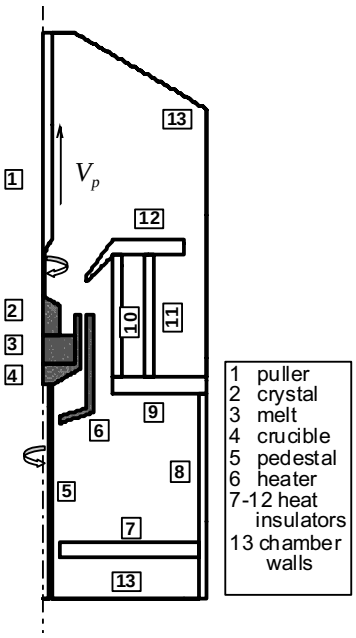


Figure 1. Configuration and domain partition of a Czochralski furnace: 3D domain (gray) and 2D domain (open).

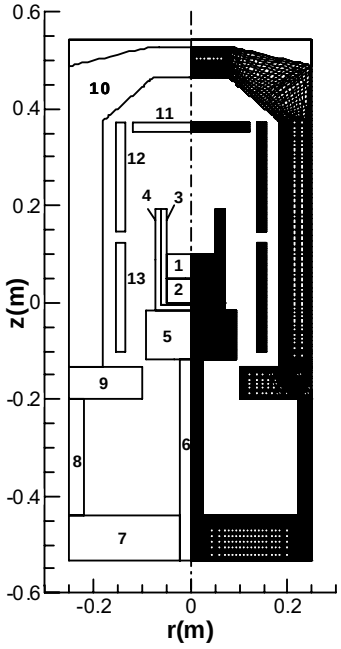


Figure 2. Configuration, domain partition (left part) and computation grid in a cross-plane (right part) of a directional solidification furnace.

For details of algorithms for an axisymmetric configuration and a nonaxisymmetric configuration, refer to Refs. (11) and (20), respectively. The governing equations for the melt convection with the assumption of incompressible laminar flow are as follows:

$$\nabla \cdot \vec{V} = 0, \quad [1]$$

$$\rho \vec{V} \cdot \nabla \vec{V} = -\nabla p + \nabla \cdot \left[\frac{1}{2} \mu \left(\nabla \vec{V} + \nabla \vec{V}^T \right) \right] - \rho \vec{g} \beta_T (T - T_m) + \vec{J} \times \vec{B}, \quad [2]$$

$$\rho c \vec{V} \cdot \nabla T = \nabla \cdot (k \nabla T), \quad [3]$$

where V, ?, p, μ, g, ß, c and k are melt velocity, melt density, melt pressure, melt viscosity, gravitational acceleration, thermal expansion coefficient, heat capacity and thermal conductivity, respectively. The melt flow is solved by a finite volume method. The global solution is obtained by an iterative procedure that consists of a set of local iterations for all block regions, calculation of radiative heat transfer in the furnace and a global conjugated iteration among them. The melt-crystal interface is calculated by a dynamic interface tracking method. Details of the model were reported elsewhere (11, 20).

Modeling and simulation of CZ-Si crystal growth

Numerical analysis of TMCZ-Si growth

A growth process was numerically simulated to demonstrate the validity of the proposed model (11). The configuration of the growth furnace is shown in Fig. 1. The furnace is in a transverse magnetic field with a non-rotating crystal and crucible. The magnetic field is applied in the x-direction (horizontal) with an intensity of 0.3 T. The crystal, melt, crucible, heater and heat insulators were included in the 3D domain. Figure 3 shows the temperature distributions in the melt as well as in solid components in the furnace.

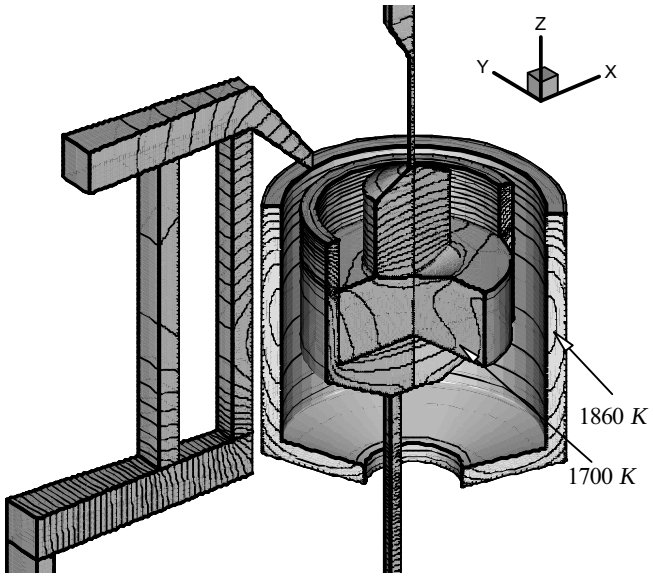


Figure 3. Temperature distribution in a CZ furnace in a transverse magnetic field. Temperature distribution in the furnace and locations under consideration. Isotherms are

plotted every 30 K in solids and every 10 K in the melt. Three-dimensional features can be clearly seen in the core region of the furnace, especially in the melt.

In Voronkov's theory (25), the ratio between the local crystal growth rate (V_g) and the temperature gradient in the crystal near the interface (G) is a key parameter of the formation of voids and interstitial clusters. Figure 4 shows results of analysis of the influence of crystal-pulling rate on distribution of the parameter V_g/G at the interface. The negative value of V_g/G means that the crystal is re-melting. The direction of the magnetic field is also indicated in the figure. From Fig. 4, we found that the distribution of the parameter V_g/G is inhomogeneous at the interface. The inhomogeneity is obvious even in the central area and the crystal is re-melted in some regions at the interface when the crystal-pulling rate is too small, as shown in Fig. 4(a) for the case with a crystal-pulling rate of 0.3 mm/min. In the case with a larger crystal-pulling rate of 1.5 mm/min, as shown in Fig. 4(b), the re-melted regions disappear and the distribution of the parameter V_g/G becomes less inhomogeneous.

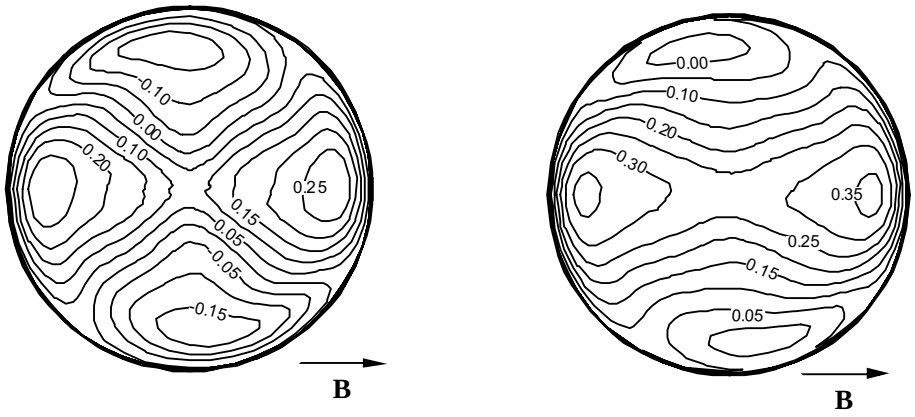


Figure 4. Influence of crystal-pulling rate on distribution of the parameter V_g/G at the melt-crystal interface. Contour lines are plotted every 0.05 $mm^2/\min K$. (a) Crystal-pulling rate of 0.3 mm/min, (b) crystal-pulling rate of 1.5 mm/min.

Figure 5 shows the axial temperature gradients in both the crystal and the melt at a melt-crystal interface as a function of crystal-pulling rate. Solid lines show the results with and without a magnetic field, while dashed lines show the results without convection in the melt, approximately corresponding to the case with a magnetic field intensity of infinite value. The arrows in the figure show the contribution of convection of the melt. The values of axial temperature gradients in the melt and crystal were obtained by averaging in the central area of the interface. The values of axial temperature gradients in the melt and crystal at the interface are not identical even when the crystal-pulling rate is zero since the thermal conductivity of the crystal is not equal to that of the melt.

These results show that the axial temperature gradients in the melt and crystal near the interface increase with increase in intensity of the magnetic field. With increase in the crystal-pulling rate, the temperature gradient near the interface in the crystal increases, while that in the melt decreases. Meanwhile, the difference becomes smaller between the case with magnetic field intensity of finite values including zero and the case without melt convection. Since this difference is due to the melt convection, this result indicates

Ex. 1016, Page 957

that the contribution of melt flow becomes smaller with increase in crystal-pulling rate.

These phenomena can be explained as follows. When we apply a magnetic field of large intensity, natural convection of the melt is suppressed, resulting in a more inhomogeneous temperature distribution in the melt. Therefore, the temperature gradient in the melt increases with increase in the magnetic field intensity. Meanwhile, because of heat balance between the liquid and solid at the interface, the temperature gradient in the crystal near the interface also increases accordingly. However, melt convection remains even if a magnetic field with a relatively large intensity is applied to the melt. As a result, even when a relatively large magnetic field of 0.3 T is applied to the system, the temperature gradients near the interface in both the crystal and the melt are far from those without melt convection, as shown in Fig. 5. On the other hand, since a larger crystal growth rate always results in lower heater power, the temperature on the sidewall of the melt becomes smaller due to the lower heater power. The temperature difference becomes smaller and the temperature distribution becomes less inhomogeneous in the melt. This in turn leads to weaker melt convection due to decrease in the thermal buoyant force induced by the temperature gradient in the melt. Therefore, with increase in the crystal-pulling rate, both the axial temperature gradient in the melt near the interface and the contribution of melt convection decrease. However, since the heat release of solidification at the interface is proportional to the crystal growth rate and it is transported away from the interface through the crystal, a larger axial temperature gradient field is generated in the crystal near the interface when the crystal-pulling rate is increased.

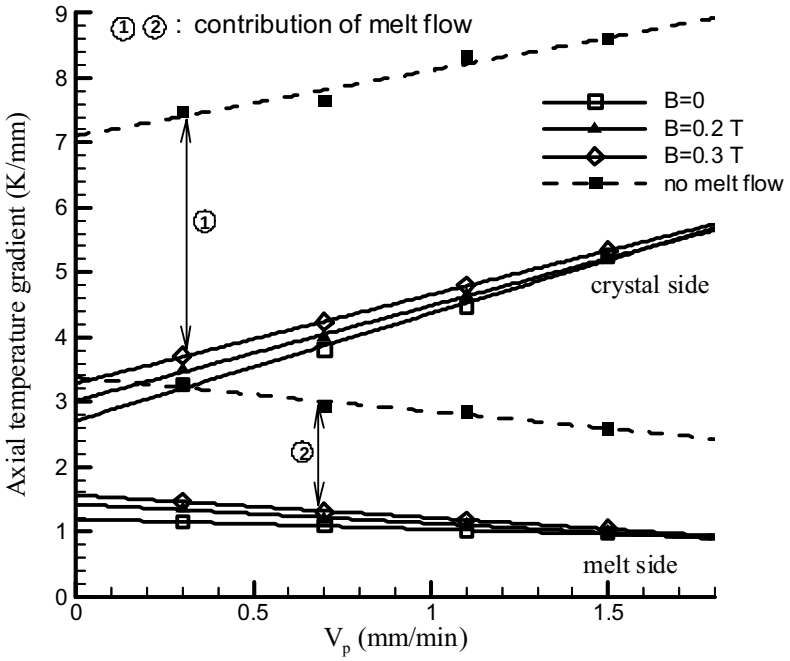


Figure 5. Axial temperature gradients in the crystal (upper part) and melt (lower part) near the interface as a function of pulling rate of a crystal at different intensities of the applied magnetic field.

Figure 6 shows interface deflection toward the melt as a function of the ratio between the crystal-pulling rate (V_p) and the temperature gradient in the crystal near the interface (G). The values of interface deflection and parameter V_p/G were obtained by averaging in the central area of the interface. The arrows in the figure show the contribution of convection in the melt. The interface moves upward to the crystal side with increase in either the magnetic field intensity or the value of the parameter V_p/G. This varying tendency is consistent with that of the axial temperature gradient in the crystal near an interface shown in Fig. 5. This is because the interface shape is mainly determined by the temperature distribution in the crystal close to the interface and the melt convection in a crucible. As explained in the previous section, when a magnetic field of large intensity is applied to the system or a larger pulling rate is applied to the crystal, the melt convection is suppressed and the axial temperature gradient in the crystal near the interface increases. The melt-crystal interface then moves upward to the crystal side in order to accommodate the increased axial temperature gradient in the crystal near the interface and the contribution of melt convection in the crucible becomes smaller.

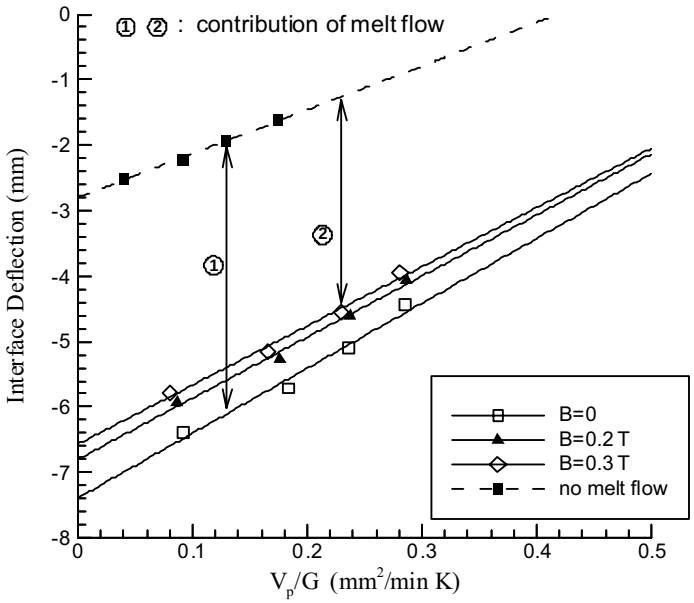


Figure 6. Interface deflection as a function of pulling rate of a crystal at different intensities of the applied magnetic field.

Effect of material property of a crucible on interface shape in a unidirectional solidification process

Figures 7 (a) and (b) show the m-c interface shape and temperature distribution of the melt and crystal obtained by 3D global analysis with two different thermal conductivities of the wall of a crucible. The m-c interface shape with high thermal conductivity shown in Fig. 7 (a) becomes more convex to the melt than that with low thermal conductivity shown in Fig. 7 (b) because the direction of heat flux at the m-c

interface with high thermal conductivity tilts toward the outside through the vertical wall of the crucible. Therefore, material property of the crucible wall is one of the key parameters for controlling the shape of a solid-liquid interface.

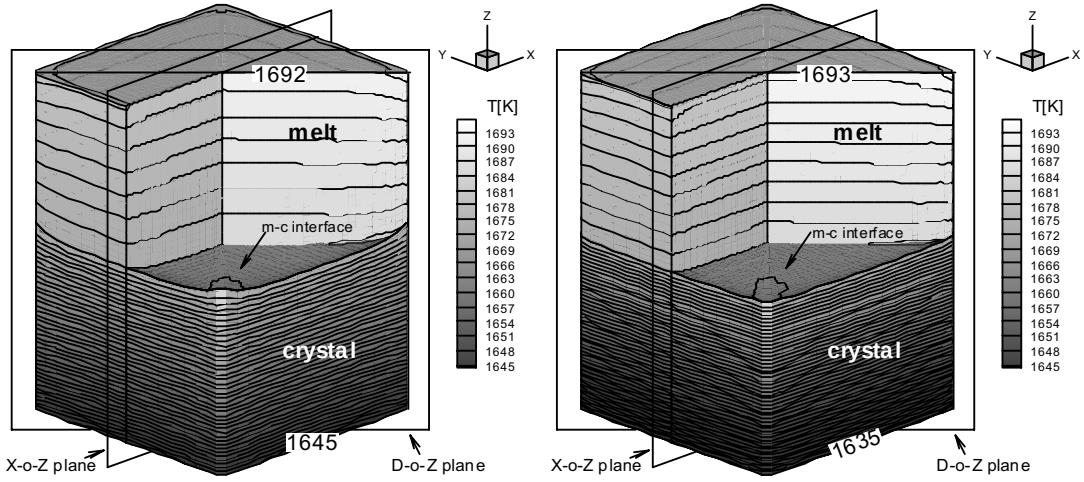


Figure 7. The m-c interface shape and temperature distribution of the melt and crystal obtained by 3D global analysis with two different thermal conductivities of the wall of a crucible: (a) 53 [W/mK] and (b) 0.267 [W/mK].

Summary

A numerical simulator was developed with a global model of heat transfer for any crystal growth taking place at high temperature. A novel 3D global model was proposed for simulation of heat transfer in any crystal growth with 3D features with moderate requirement of computer resources. The application of this numerical simulator to a CZ growth and a directional solidification process for Si crystals was introduced. Some typical results were presented, and the results show the importance and effectiveness of numerical simulation in analyzing and improving these kinds of Si crystal growth processes.

Some ongoing researches to improve the simulator include: 1 Modeling of thermal stress and dislocation defect evolution in a crystal growth of silicon; 2 Modeling of impurities precipitation; 3 Turbulence modeling of melt flow with large volume in case of growing large-size crystals.

Acknowledgements

This work was supported by a NEDO project, a Grant-in-Aid for Scientific Research (B) 19360012 from the Japanese Ministry of Education, Science, Sports and Culture.

References

1. N. Kobayashi, *J. Crystal Growth*, **43**, 3573 (1978).
2. A. Krauze A. Muiznieks, A. Muhlbauer, Th. Wetzel, W. v. Ammon, *J. Crystal Growth*, **262**, 157 (2004).

Ex. 1016, Page 960

3. K. Kakimoto, L.J. Liu, *J. Crystal Res. Technol*., **38**, 716 (2003).
4. J. J. Derby, R.A. Brown, *J. Crystal Growth*, **74**, 605 (1968).
5. L. J. Atherton, J. J. Derby, R. A. Brown, *J. Crystal Growth*, **84**, 57 (1987).
6. F. Dupret, P. Nicodeme, Y. Ryckmans, P. Wouters, M. J.Crochet, *Int. J. Heat Mass Transfer*, **33**, 1849 (1990).
7. M. Li, Y. R. Li, N. Imaishi, T. Tsukada, *J. Crystal Growth*, **234**, 32 (2002).
8. V. V. Kalaev, I. Evstratov, Y. Makarov, *J. Crystal Growth*, **249**, 87 (2003).
9. D. Vizman, S. Eichler, J. Friedrich, G. Müller, *J. Crystal Growth*, **266**, 396 (2004).
10. V. V. Kalaev, D. P. Lukanin, *J. Crystal Growth*, **250**, 203 (2003).
11. L. J. Liu, K. Kakimoto, *Int. J. Heat Mass Transfer*, **48**, 4481 (2005).
12. L. J. Liu, K. Kakimoto, *Int. J. Heat Mass Transfer*, **48**, 4492 (2005).
13. K. Kakimoto, L. Liu, H. Miyazawa, S. Nakano, D. Kashiwagi, X. J. Chen, and Y. Kangawa. Crystal Research & Technol., **42**, 1185 (2007).
14. D. Kashiwagi, R. Gejo, Y. Kangawa, L. Liu, F. Kawamura, Y. Mori, T. Sasaki and K. Kakimoto, *Journal of Crystal Growth*, **310**, 1790 (2008).
15. X. J. Chen, L. J. Liu, H. Tezuka, A. Usuki, K. Kakimoto, *Crystal Research and Technol.*, **42**, 971 (2007).
16. X. J. Chen, L. J. Liu, H. Tezuka, Y. Usuki, K. Kakimoto, *J. Crystal Growth*, **310**, 1810 (2008).
17. L. J. Liu, K. Kakimoto, *Cryst. Res. Technol.*, **40**, 347 (2005).
18. L. J. Liu, S. Nakano, K. Kakimoto, *J. Crystal Growth*, **282**, 49 (2005).
19. L. J. Liu, S. Nakano, K. Kakimoto, *J. Crystal Growth*, **299**, 48, (2007).
20. L. J. Liu, S. Nakano and K. Kakimoto, *J. Crystal Growth*, **303**, 165 (2007).
21. L.J. Liu, S. Nakano and K. Kakimoto, *J. Crystal Growth*, **292**, 515 (2006).
22. L. J. Liu, S. Nakano and K. Kakimoto, *J. Crystal Growth*, **310**, 2192 (2008).
23. H. Miyazawa, L.J. Liu, S. Hisamatsu and K. Kakimoto, *J. Crystal Growth*, **310**, 1034 (2008).
24. T. Kajigaya, T. Kimura, Y. Kadota, *J. Crystal Growth*, **112**, 123 (1991).
25. V. V. Voronkov, *J. Crystal Growth*, **59**, 625 (1982).
26. H. Hirata, K. Hoshikawa, *J. Jpn. Assoc. Crystal Growth*, **15**, 207 (1988).

ECS Transactions, 18 (1) 935-943 (2009)
10.1149/1.3096557 © The Electrochemical Society

Global Simulation of CZ and FZ Bulk Crystal Growth: from Quasi-Dynamic and Dynamic Modelling to Process Control and Crystal Quality Optimization

F. Dupret[a,b], R. Rolinski[b], L. Wu[b], F. Loix[b],
A. de Potter[b], N. Van den Bogaert[b], and V. Regnier[b]

[a] CESAME Research Center, Université catholique de Louvain,
B-1348 Louvain-la-Neuve, Belgium
[b] FEMAG*Soft* S.A. Company, B-1435 Mont-Saint-Guibert, Belgium

When Silicon ingots are grown for IC or PV applications, the crystal diameter has to conform to current market requirements while the crystal defects and composition have to be perfectly controlled. The present paper is devoted to illustrate CZ and FZ process optimization by use of the time-dependent FEMAG.2 software. First, the "off-line control" technique is devoted to determine the heater power and pull rate histories required to obtain a crystal of constant diameter and optimal quality by means of fully transient simulations. Other command parameters such as crystal and crucible rotation rates, magnetic field intensity, etc., can be optimized as well. Secondly, the use of quasi-dynamic numerical simulations shows very efficient to study the transient effects experienced by the system during bulk crystal growth, without necessitating a fully dynamic simulation. A main application of this technique is to improve the design of an automatic controller of the growth process.

Introduction

The growth of Silicon (Si) crystals by the Czochralski (Cz) or Floating Zone (FZ) technique for electronic (IC) applications has been governed for more than 50 years by two technological objectives. First, the crystal diameter has to be nearly constant and the largest possible according to market requirements. Secondly, the product quality has to be perfectly controlled in terms of crystal defects and composition. On the other hand, growing Cz Si crystals for photo-voltaic (PV) applications requires to minimize both the energy consumption and the growth duration without, however, generating a too large content of micro-voids in the crystal. Achieving these somewhat contradictory goals is by no means easy since increasing the crystal diameter requires a larger melt volume and hence results in a much more complex melt flow regime with complicated heat, momentum and species transport effects, a very sensitive solid-liquid interface shape (with a less uniform thermal gradient), and in general an enhanced dynamic system behavior. A similar enhancement of the system dynamic behavior can result from the use of a high pull rate to increase the growth speed. Therefore, in general, designing the furnace hot zone requires to introduce appropriate heat shields in order to well-control the radiation heat transfer while in Cz growth a satisfactory melt flow pattern can only be obtained for large diameter crystals by the action of transverse or configured magnetic fields. Moreover the selection of optimal process parameters (heater power or inductor voltage, crystal pulling rate, crystal and crucible or feed-rod rotation rates, magnetic field intensity if any, ambient gas flow rate, etc.) becomes much more difficult in view of the

increased system nonlinearity and time-dependency, especially during the critical process stages (necking, shouldering, tail-end stage, crystal detachment, …).

Nonetheless, in front of these technological difficulties, huge progress has been achieved in the last decades in different scientific domains. First, the physics of radiation and convection, on the one hand, and of defect formation and transport in Si crystals, on the other hand, is much better known, and hence the mathematical models governing Si growth are better and better established. Whereas important improvements remain necessary in the modeling of turbulence in the melt and the ambient gas (including the modeling of melt turbulence under the effect of a magnetic field) while the material parameters governing point- and micro-defect evolution in Si single crystals still are insufficiently known, an almost complete picture of the physics of Si growth today is available. Secondly, numerical methods and computers have also quickly progressed since the development of the first models of Cz and FZ growth achieved in the 1980's. Nowadays the quasi-steady or time-dependent simulation of these processes has become possible in an acceptable computing time, with sufficiently refined meshes to resolve the key details of the problem, and with appropriate numerical techniques to handle the system deforming geometry (which comprises several moving components together with free boundaries such as the melt-crystal and melt-gas, and possibly the melt-feed rod, interfaces).

Having at one's disposal the appropriate physical models, numerical tools and computer hardware directly opens the route to Cz or FZ process optimization by means of numerical simulation. The objective of the present paper is to illustrate how this straightforward strategy can be applied by use of the FEMAG-CZ or FEMAG-FZ software (collectively called FEMAG) as today co-developed by FEMAG*Soft* S.A. Company and the CESAME research center of the Université de Louvain (Belgium).

The global and time-dependent FEMAG-CZ and FEMAG-FZ models

FEMAG software products are currently used by major Si growers in the world. The numerical model is both global and dynamic, and takes the effect of melt convection into account. Diffuse surface radiation is considered, while frequency-dependent material properties are modeled by means of the band-energy technique. Radiation calculations are performed by accurate integration of the surface view factors around the furnace axis. The global model of FEMAG-CZ takes all important physical phenomena of the growth process into account, i.e. heat transfer by conduction in every furnace constituent and by convection in the melt and gas phases, together with solidification at the solid/liquid interface, surface tension, etc. Geometrical unknowns are dynamically coupled to the other unknowns, i.e. temperature field, velocity field, electrical potential (in the presence of a magnetic field), etc., leading to a complex non-linear system of equations whose solution is found by use of a decoupled scheme at every time step of the simulation.

Whereas in its first generation the FEMAG.1 software family already performed global quasi-steady or time-dependent simulations, the applications were restricted to top cone, shouldering and body growth stages. Both laminar and non-laminar flow models were considered, including or not the effect of axisymmetric (axial, cusp or rotating) magnetic fields. The objective of launching the FEMAG.2 software generation has been

to provide a fully automatic dynamic simulator predicting the entire growth process, without any manual user's intervention during simulation, while handling correctly the switches between the growth stages (poly-crystal melting, seeding, conical growth, shouldering, body growth, tail-end stage and after growth cooling), together with coupling the dynamic calculations with accurate melt flow prediction.

However a significant difficulty lay in the important evolution of the system geometry during a complete growth process. Indeed, the solidified region is very small during seeding and subsequently becomes larger and larger, while in Cz growth, the volume of the molten region decreases continually and can take a complex shape during tail-end stage. Similar geometrical difficulties are encountered in the modeling of FZ growth. The solution adopted in FEMAG.2 combines several approaches based on a representation of the furnace by means of deforming unstructured meshes together with automatic mesh generation. New geometrical methods have been designed to allow easy calculation of the different system free surfaces (solidification front, melt/gas interface including crystal/melt and crucible/melt or feed rod/melt menisci, and crystal/gas surface) through a single mathematical formulation, valid for all possible configurations. These methods allow the user to perform easy time-dependent simulations even for those stages of the process where important geometrical changes occur.

Another important issue to address in the development of FEMAG.2 was related to the complexity of dynamic melt flow modeling. In fact, in Cz Si growth, two time scales have to be considered. The short scale, typically of the order of tens of seconds, governs the basic transients associated with melt flow oscillations at the different process stages – noting that the flow exhibits weak turbulence superposed to 3D azimuthal and temporal structured oscillations. The long time scale, typically of the order of tens of minutes, is associated with the flow and heat transfer transients caused by the long term system evolution – since melt and crystal shapes continually change and can be very complex during some growth stages (especially during the beginning and the end of the growth process), while the heat transfer is strongly affected by melt-crystal interface deformation and by the heater power evolution required to grow a crystal of the prescribed shape. Although the melt flow in FZ Si growth is normally laminar, its behavior is also oscillatory and rather complex.

Therefore the basic problem to address in order to couple flow predictions with global dynamic simulations was to develop a model able to average the effect of the flow oscillations above the steady laminar regime. To this end, two models (viz. the "Mixing Length" and k-l models) were investigated, as based on the Boussinesq approach with use of an additional eddy viscosity and thermal conductivity to account for the enhancing effect of the flow oscillations on heat and momentum transfer in the melt or the ambient gas. In the former case the additional viscosity and thermal conductivity are functions of the mixing length l only, while they are functions of the mixing length and the additional kinetic energy k in the latter case. In both cases, the mixing length represents a key parameter whose accurate modeling allows the user to fit the numerical predictions to experimental measurements, by means of an appropriate tuning of the boundary layer thicknesses along the melt or gas interfaces with the surroundings.

A typical time-dependent inverse global result obtained by means of FEMAG-CZ.2 is shown in Fig. 1.

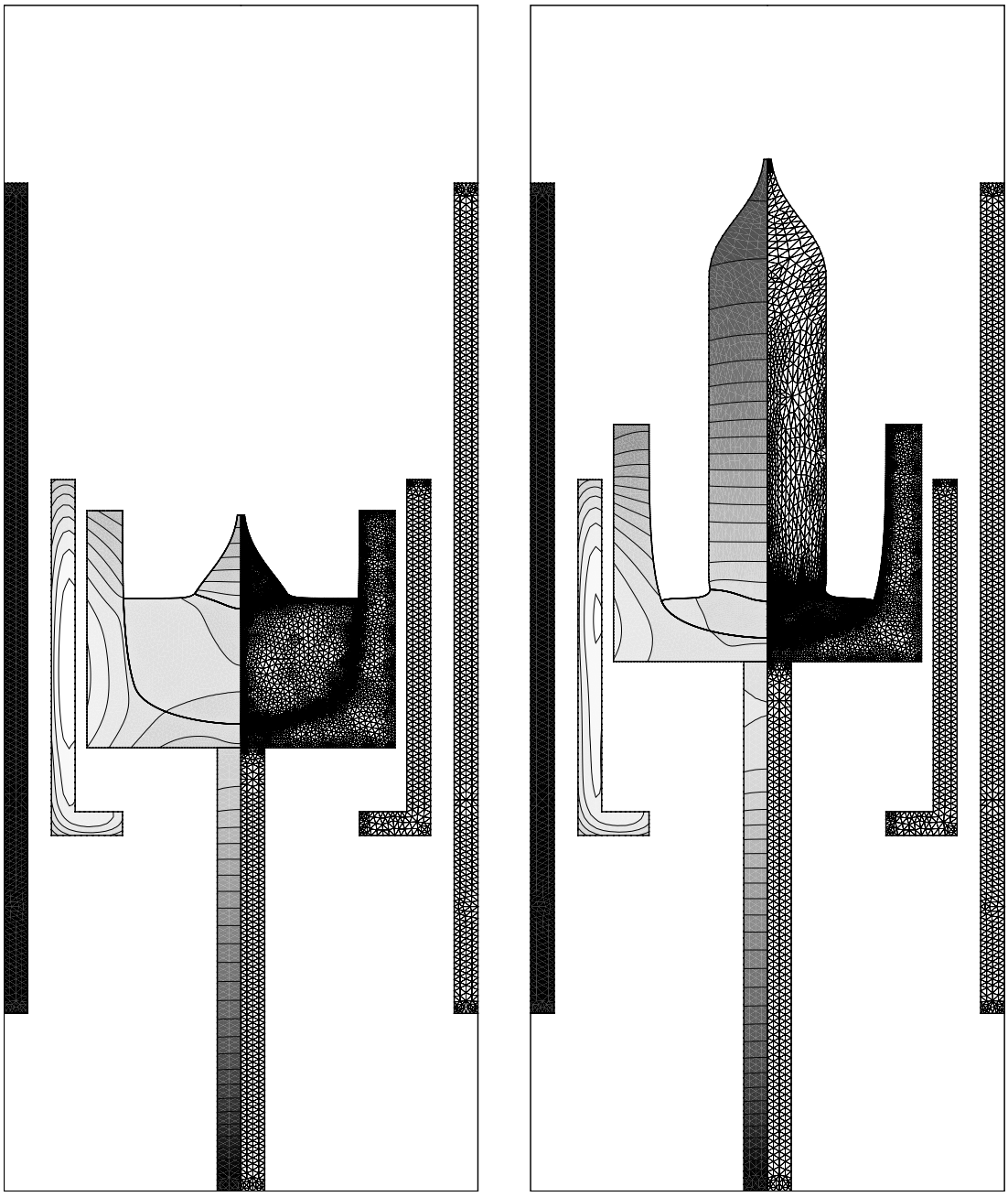


Figure 1. Growth of a 100 mm Si crystal (pulling rate: 0.5 mm/min; crystal rotation rate: 10 rpm; crucible rotation rate: -5 rpm). Temperature field by steps of 50 K and deforming mesh calculated at 2 growth stages by means of a transient FEMAG-CZ.2 simulation.

Prediction of defects and species distribution in the solid phase

A key issue in Cz or FZ Si growth is to reduce defect density to the lowest possible level. Therefore the effort paid for several years to understand the mechanisms governing the

formation and evolution of vacancies, self-interstitials, micro-voids, clusters, dislocation loops, etc., in the growing crystal has brought reasonably good models on the basis of Voronkov's theory. However numerical experiments reveal a high sensitivity of the defect distribution both to the material parameters governing defect diffusion, recombination and agglomeration, and to the thermal gradient prevailing in the crystal (especially along the solidification front). Therefore accurate numerical models are needed when the objective is to improve the growth process.

Using the FEMAG.2 software, we here present a fully time-dependent model devoted to predict the global heat transfer in the furnace, the solid-liquid interface shape, and the resulting distributions of point- and micro-defects in the crystal as calculated from the Sinno-Dornberger (S-D) model together with an extension of the lumped model of Voronkov and Kulkarni. All the transients are considered including the effects of crystal and crucible lift, of the heat capacities of the furnace constituents, of the thermal inertia of the solidification front, and of the dynamic defect governing laws. In addition to the classical point-defect evolution mechanisms, a new lumped model is introduced to calculate the formation and growth of micro-defects in order to predict their densities and size distributions anywhere in the crystal.

We also show that dynamic effects deeply affect the defect distribution in the crystal. Indeed, the interface deformation caused by any change of the operating conditions directly affects the temperature gradient above the interface and the resulting defect densities through the V/G ratio. Moreover, as point- and micro-defects are transported while diffusing and reacting, the defect distribution in the crystal is a picture of the past history of defect generation and can only be predicted by means of a dynamic model.

Another key issue in Si growth is to control the density of oxygen (in Cz growth) and any other species (including dopants and impurities) inside the crystal. To address this problem several important difficulties have to be solved. First, the transport of any species in the melt is governed by complex mechanisms where the flow oscillations play a major role with respect to the very low molecular diffusivity. Here again an additional diffusivity has to be introduced in order to account for the enhancing effect of the flow oscillations on mass transfer in the melt (or the ambient gas), and the Mixing Length or *k-l* model can readily be used for this purpose. The second and more important difficulty to overcome in order to simulate species transport in the melt and the gas in Cz or FZ Si growth is to well-model the boundary conditions along the different interfaces separating the flows from their surroundings. In particular the way oxygen escapes from the melt to the gas in Cz growth strongly affects its distribution in the crystal and therefore the boundary layers along both sides of the melt-gas interface have to be accurately modeled. On the other hand segregation can deeply affect the distribution of some dopants in Si crystals. Here again addressing this problem requires to appropriately model the melt boundary layer located along the solidification interface.

The ultimate goal of these investigations is to get a complete and reliable picture of the entire set of mechanisms governing the quality of Cz- or FZ-grown Si crystals.

Ex. 1016, Page 967

Off-line and on-line process control

The main interest of a time-dependent simulator such as FEMAG.2 is to help the crystal grower in the design of a new furnace (including the design of the furnace hot zone) and the optimization of the growth process. This latter objective can be achieved in two ways. On the one hand, the "off-line" control technique is devoted to optimize the selection of the crystal growth processing conditions. In other words, off-line control provides master (or profile) curves that have to be followed by the processing conditions in order to grow a crystal of the highest quality. On the other hand, various kinds of time-dependent simulations can be very efficiently used to improve the design of a Cz or FZ "on-line" (or off-line) controller.

Off-line control

In general dynamic simulations can be direct or inverse. In this latter case, some constraints are introduced (such as prescribing the crystal shape) and an equal number of natural inputs (heater power or inductor voltage, …) are calculated to satisfy these constraints. Typically, in inverse Cz simulations, the heater power history is calculated to constrain the crystal to grow with the prescribed shape. Off-line control consists in extending the inverse simulation technique in order to optimize the growth process.

Precisely, off-line control aims at determining an appropriate evolution of the processing parameters (heater power; pulling rate; crystal, crucible or feed rod rotation rates; magnetic field intensity, …) in order to optimize crystal shape and quality. Here, the simulation is managed by a controller, which retroacts by searching optimal command parameters to satisfy selected quality criteria. Therefore the simulator plays the role of the real process, while the off-line controller supervises the simulation. At each time step the retroaction loop transmits information about the system state to the controller, which in turn evaluates the required evolution of the processing parameters in order to optimize crystal quality.

Typically, in Si growth, the objective is to determine the heater power and pull rate histories required to obtain a constant diameter crystal of optimal quality. Additional command parameters such as crystal and crucible or feed rod rotation rates, magnetic field intensity if any, etc., can be optimized as well. Crystal quality is measured by simulated results such as the defect density above the crystal-melt interface, or the deflection of this interface, etc.

On-line control

The other significant interest of time-dependent simulations is to provide various attractive techniques to improve the process control. High quality controller design can be achieved by means of quasi-dynamic and direct dynamic simulations.

Quasi-dynamic simulations. For a given crystal shape and vertical position, a quasi-steady global model of the pulling process is defined by making the assumption that all operating conditions are kept constant and that the solid-liquid interface shape is kept fixed while the solidification heat is released from the solid-liquid interface in proportion to the pull rate. The resulting equations do not contain transient terms anymore. This

model is inverse, because it is necessary to let one of the process parameters (typically the heater power) become an unknown whose value is calculated in order to get the prescribed crystal diameter.

The *quasi-dynamic* simulation technique is an extension of the quasi-steady model where the crystal position is again frozen together with the geometry (including the crucible height), while the melt-crystal interface is allowed to dynamically deform and the other process parameters are allowed to change with time. In other words, the geometrical modifications resulting from the upward motion of the crystal, the crucible, the pedestal, etc., are not considered, while the sole effect of the crystal vertical motion is to release solidification heat along the melt-crystal interface. However, this interface itself (and only this interface) is allowed to deform, hence releasing some extra-solidification heat, and the transients of the melt and all other components are taken into account together with any possible change of the operating conditions (pulling rate, rotation rates, etc.). In this way the simulation is able to capture the short time scales governing the melt flow dynamics and their effect on the melt-crystal interface deformation and the heat transfer in the crystal just above this interface. Obviously the quasi-dynamic simulation model represents a key tool to improve the design of an off-line or an on-line controller of the growth process.

As an example, Figs. 2-5 depict the results obtained by designing a controller on the basis of quasi-dynamic simulations. The objective is to optimize the quality of an Si crystal as measured either by the axial interface deflection z_{ax} (defined as the difference between the vertical positions of the melt-crystal interface at the tri-junction and on the axis) or by the well-known v/G ratio. The control variable is the pulling rate v_{pul}. The short time constants governing the system have been identified by performing a quasi-dynamic simulation of the system response to a 5% stepwise decrease of the pulling rate. A simple PI controller was subsequently designed from these results and inverse, off-line controlled, dynamic simulations were performed to analyze the effect of this controller on the resulting crystal quality.

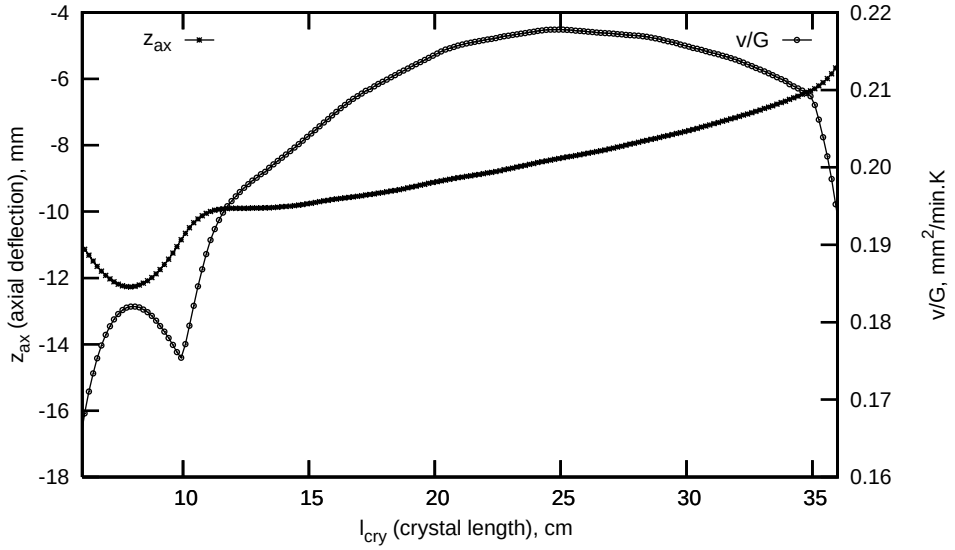


Figure 2. Uncontrolled simulation of the problem of Fig. 1. Evolution of the axial interface deflection z_{ax} and the v/G ratio as a function of the grown crystal length l_{cry}.

Ex. 1016, Page 969

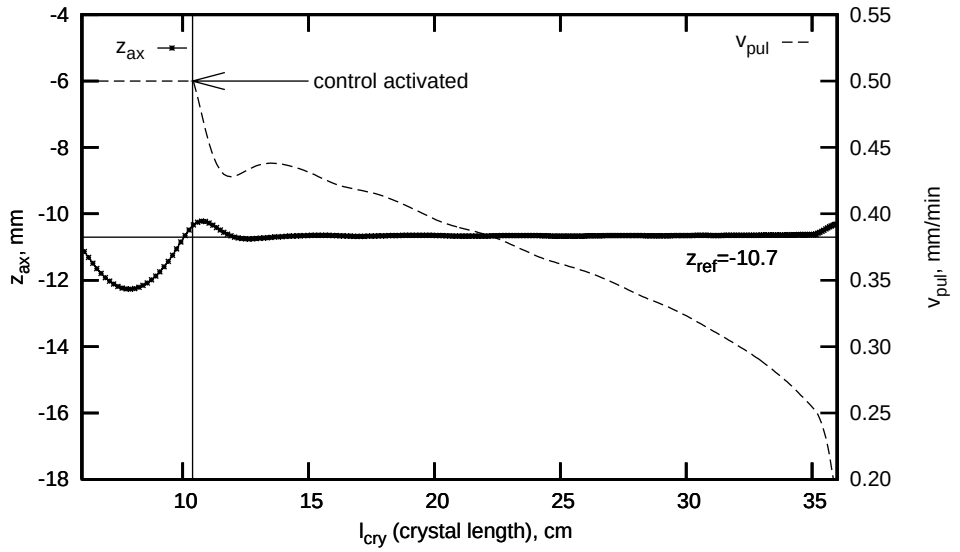


Figure 3. First transient simulation, with an off-line controlled axial interface deflection z_{ax} while the pulling rate v_{pul} is the control variable.

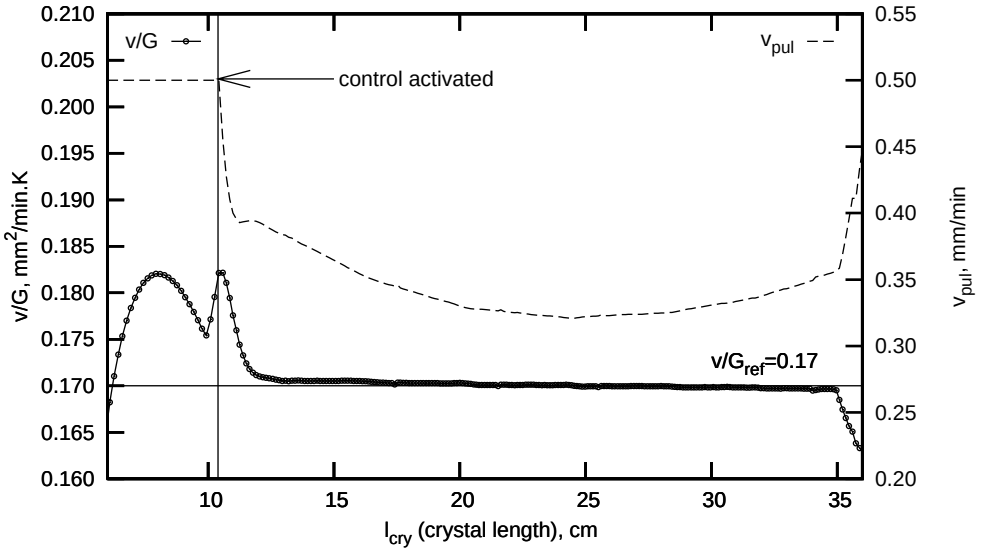


Figure 4. Second transient simulation, with an off-line controlled v/G ratio while the pulling rate v_{pul} is the control variable.

The results clearly show that our procedure was quite successful. The quality of the identified controller can immediately be observed from Figs. 3 and 4, where the system quickly stabilizes to the prescribed z_{ax} or v/G by appropriate action on the pulling rate. Moreover, Fig. 5 indicates in addition that well-controlling either of these variables should result in an improved crystal quality as long as the calculated difference between the interstitial and vacancy densities C_I and C_V represents a good indicator of this quality (it is worth recalling here that growing Si crystals with a low C_I - C_V is a requirement for several IC components).

Ex. 1016, Page 970

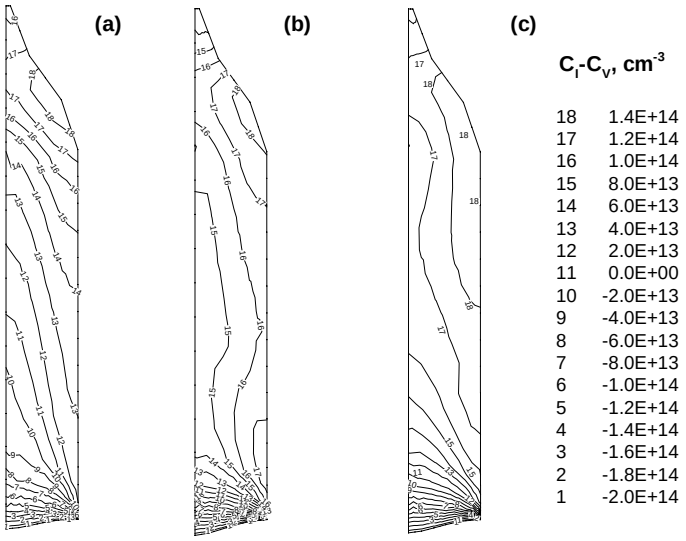


Figure 5. Isovalues of C_I - C_V in the crystal for the three considered transient simulations of the problem of Fig. 1. (a) uncontrolled simulation; (b) off-line controlled axial interface deflection z_{ax}; (c) off-line controlled *v/G* ratio.

Direct dynamic simulations. Finally, the *direct dynamic* simulation technique is defined by first performing an inverse dynamic reference simulation (where the pull rate profile and the crystal shape are imposed while the heater power history is calculated to obtain the imposed crystal shape), and then by perturbing this reference simulation by slightly modifying the heater power or the pull rate history and calculating the resulting crystal shape. Typically, stepwise decreases of the heater power or the pulling rate profiles can be applied and, in this way, the direct dynamic simulations are able to capture the longer time scales of the growth furnace dynamics and their delayed effect on the crystal diameter. Obviously this technique represents a valuable complementary tool for the design of a crystal growth controller with respect to quasi-dynamic simulations.

References

1. V.V. Voronkov, *J. Crystal Growth*, **59**, 625, (1982).
2. F. Dupret, P. Nicodème, Y. Ryckmans, P. Wouters and M.J. Crochet, *Int. J. Heat Mass Transfer*, **33**, 1849, (1990).
3. F. Dupret and N. Van den Bogaert, in *Handbook of Crystal Growth*, vol. 2b, ch. 15, D.T.J. Hurle, Editor, p. 875, North-Holland (1994).
4. R. Assaker, N. Van den Bogaert and F. Dupret, *Magnetohydrodynamics*, **31**, 254, (1995).
5. N. Van den Bogaert and F. Dupret, *J. Crystal Growth*, **166**, 446, (1996); **171**, 65, (1997); **171**, 77, (1997).
6. R. Assaker, N. Van den Bogaert and F. Dupret, *J. Crystal Growth*, **180**, 450, (1997).
7. T. Sinno, E. Dornberger, R.A. Brown, W. von Ammon and F. Dupret, *Materials Science and Engineering: R Reports*, **28**, 149, (2000).
8. M.S. Kulkarni, V.V. Voronkov and R. Falster, *J. Electrochem. Soc.*, **151**, G663, (2004).
9. M.S. Kulkarni and V.V. Voronkov, *J. Electrochem. Soc.*, **152**, G781, (2005).

Ex. 1016, Page 971

ECS Transactions, 18 (1) 945-957 (2009)
10.1149/1.3096558 © The Electrochemical Society

Agglomeration of Vacancies into Voids and Oxide Particles in Silicon Crystals

V. V. Voronkov
MEMC Electronic Materials, Merano 39012, Italy

The vacancies, incorporated into a growing crystal, agglomerate by two parallel paths - producing voids and oxide particles. The voids dominate if the vacancy concentration C_0 is high while the oxide particles dominate at lower C_0. The void and particle properties are simulated assuming that the particles are thin plates. The initial single-octahedral void shape can later change to double-octahedral, due to oxygen film adsorbed at the void facets. By a similar mechanism, platelet shape is developed in nitrogen-doped crystals.

Introduction

Modern Czochralski (CZ) silicon crystals are normally grown in the vacancy mode: after the initial stage of annihilation of vacancies (V) and self-interstitials (I), the remaining defects are vacancies. This mode is achieved when the ratio V/G of the growth rate V to the axial temperature gradient G exceeds some critical value, $(V/G)_{cr}$ (1). The reported numbers for the critical ratio are scattered around 0.18 mm^2/minK (1-5). The incorporated vacancy concentration C_0 increases, upon increasing V/G (1, 6), from zero at $V/G = (V/G)_{cr}$, to the saturated value $C_S = C_{Ve} - C_{Ie}$ - the difference between the equilibrium concentrations of vacancies and self-interstitials at the melting point (T_m). By tracing the total void volume in dependence of V/G (4), C_S is estimated to be about $1.6x10^{14}$ cm^{-3}.

In the course of crystal cooling, the vacancy supersaturation increases, and at some moment becomes sufficient for vacancy clustering. The preferable form of clusters is expected to be octahedral voids (1), and indeed these microdefects were observed by Transmission Electron Microscopy (TEM) (7-9). At lower C_0, the preferable path of clustering should switch to joint agglomeration of vacancies and oxygen into oxide particles (10). Accordingly, a narrow band of oxide particles (P-band) is formed at the periphery of the vacancy region – where the local value of C_0 is low. In the V-mode, this band is close to the crystal surface. If a crystal is grown in a mixed mode - and contains both V-region and I-region separated by a sharp V/I boundary - the P-band is located close to the V/I boundary. A well known phenomenon of a narrow OSF-ring – a ring of Stacking Faults (SF) in Oxidized wafers (11-13) – is caused by the particles of the P-band that serve as sites for aggregation of oxidation-induced self-interstitials into SFs. The P-band is thus revealed as an OSF-ring.

A quantitative modeling of void and oxide particle formation, in dependence of the three basic parameters - which are C_0, the cooling rate through the relevant temperature interval of clustering, and the oxygen concentration C_{ox} - was performed (10, 14, 15) but some problems remain not yet resolved. One is the way of void production: either by direct nucleation (10) or by initial nucleation of amorphous oxide particles (14). The amorphous oxide becomes severely stretched by supply of vacancies which leads to cavitation: a small cavity (bubble) is nucleated within oxide, and the defect then evolves into a regular void. Some particles remain non-cavitated, and thus both voids and particles are produced sequentially, by a single nucleation path. The cavitation scenario has some advantages but at the same time some drawbacks, and in the present work we

stick to the "classical" scenario when voids and oxide particles are produced in parallel, by two independent nucleation paths.

In the previous simulations (10, 14, 15), the oxide particles were assumed to be globular (spherical or octahedral); the predicted particle density, in the middle of the P-band, was over 10^9 cm^{-3} – while the experimental density was essentially lower, about 10^8 cm^{-3} (3, 13). It is known, that in the course of wafer annealing at around 750°C, the oxide particles initially exist as plates of monoatomic thickness that can be revealed only by high-resolution TEM (16); later they are transformed into strained platelets of amorphous oxide, revealed by conventional TEM and known as the major oxygen precipitates in this temperature range. It is natural to assume that the oxide particles nucleated during crystal growth are the same plates (so called "ninja" precipitates (17)).

An important phenomenon in void and particle formation is vacancy binding by oxygen, into VO and VO_2 complexes (10, 18). There is a characteristic binding temperature T_b that separates a higher-T range (where the dominant vacancy state is V, free vacancy) and a lower-T range (where the dominant state is VO_2, bound vacancy). Tentatively, T_b was assumed to be 1050°C – well below the void nucleation stage that normally occurs at around 1100°C. Recently there appeared some evidence in favor of a higher value of T_b – close to 1100°C (this issue is to be discussed elsewhere). It means that voids are formed when the vacancy binding is already appreciable.

Accordingly, in the present work the simulations of void and particle formation were performed using two major modifications: (1) the oxide particles are thin plates (monolayers of oxygen); (2) the binding temperature T_b is 1100°C (for a standard oxygen concentration of 14 ppma).

Another important issue to be discussed is the void morphology. Depending on the growth parameters, the voids were reported (7-9, 19, 20) to be either single-octahedral or double-octahedral structure (of two adjacent octahedra) – or sometimes even a triple structure. The void morphology is further changed – to become of a platelet type - in nitrogen-doped CZ crystals (21). These morphological peculiarities are explained by formation of adsorbed impurity films at the void surface.

Nucleation of voids and oxide plates

In the nucleation problem, the most important part is the nucleation barrier. The free energy change, due to formation of a vacancy cluster (a small void), in dependence of the number m of agglomerated vacancies, is $F(m) = -m\,\mu_V + F_{sur}(m)$. Here $\mu_V = kT \log(C_V/C_{Ve})$ is the vacancy chemical potential (C_V is the actual vacancy concentration and $C_{Ve}(T)$ is the equilibrium vacancy concentration). The surface energy of a void, F_{sur}, is approximated by a smooth function proportional to the surface area and thus represented as $\lambda\, m^{2/3}$ (where λ is the surface energy coefficient). The void nucleation barrier – the highest value of $F(m)$ - is then $F_v = (4/27)\,\lambda^3 / \mu_V^2$. The void nucleation rate is known to be sensitive to the oxygen content C_{ox} (22, 23) which implies that a void contains oxygen atoms (apparently, in adsorbed state). It is assumed that a continuous oxygen film (monolayer) is present at the (111) facets of the void; each oxygen atom is attached to two silicon dangling bonds. The coefficient λ is then dependent on the oxygen chemical potential $\mu_{ox} = kT \log(C_{ox} / C_e)$ where C_e is the oxygen solubility (from a bulk non-strained dioxide). The void nucleation barrier is then modified (24) to be

$$F_v = (2/3)\,(u - \mu_{ox})^3 / \mu_V^2 \quad . \qquad [1]$$

The energy u is a fitting parameter to be specified later.

The nucleation rate of voids includes the rate-limiting attachment frequency ω (probability of an attachment of relevant units per unit time). For voids without oxygen film - which is the case for Float-Zoned crystals - ω refers to vacancies; it is equal to $4 \pi R_v^* D_V C_V$, for diffusion-limited attachment. Here R_v^* is the void critical radius and D_V is the vacancy diffusivity. A typical value of ω is 3×10^3 s^{-1}. For oxygen-covered voids the size increment is limited rather by oxygen attachment, and one should use the expression $\omega = 4 \pi R_v^* D_{ox} C_{ox}$ where D_{ox} is the oxygen diffusivity; then a typical value of ω, at 1100°C, is 30 s^{-1}. The void nucleation rate is expressed, by the classical nucleation theory

$$I_v = \omega Z [\rho \exp(-F_v / kT)] \quad . \qquad [2]$$

The term in square brackets represents the equilibrium concentration of the critical void nuclei (ρ is the lattice site density) and Z is the Zeldovich factor. One do not need to be too scrupulous in defining the pre-exponential factor in equation [2] because of a huge sensitivity of the barrier factor, $\exp(-F_v/kT)$, to the energy parameter u. A change in the prefactor (even by some orders of magnitude) is compensated by a tiny change in u. For this reason the expression [2] can be safely simplified by putting $Z = 0.1$ (a typical value of this factor) and replacing the critical radius R_v^* with a constant representative value, like $R^* = 0.3$ nm. Alternatively one can use the equation obtained from [2] by substituting precise expressions for Z and R_v^* (25).

Oxide volume, per one Si atom, is about twice as large as that for the matrix; to avoid the strain, about 1 vacancy should be added for every 2 oxygen atoms. A coherent plate should be composed of vacancies and oxygen atoms in a strict proportion 1:2. Formally a plate consists of VO_2 units (although physically it is built by arriving vacancies and oxygen atoms). Such a structure was discussed in (26). The free energy gain per one unit is $\mu_p = \mu_V + 2 \mu_{ox} - f_s$ where f_s is the contribution from the two closely-spaced coherent oxide/matrix interfaces. The total free energy $F(m)$, in dependence of the number m of units, is $- m \mu_p + F_{edge}$. The last term is due to the plate edge; it is proportional to the edge length and thus represented as $\lambda_p m^{1/2}$. The plate nucleation barrier is

$$F_p = (1/4) \lambda_p^2 / (\mu_V + 2 \mu_{ox} - f_s) \quad . \qquad [3]$$

The oxygen solubility from the plate, C_{ep}, defined by condition $2 \mu_{ox} - f_s = 0$, exceeds the normal solubility C_e by a factor $\exp(f_s/2kT)$. The value of C_{ep} for ninja plates was estimated as 9 ppma at 1000°C (17). The solubility $C_e(T)$ is taken after (27, 28); it is defined by the value 2×10^{18} cm^{-3} (40 ppma) at T_m and the activation energy 1.5 eV. At 1000°C, $C_e = 1.4$ ppma. Accordingly, $f_s = 2 kT \log(C_{pe}/C_e) = 0.4$ eV. The other nucleation parameter for plates, the edge energy coefficient λ_p, is a fitting parameter specified later.

The plate nucleation rate I_p is defined by the equation similar to [2]:

$$I_p = \omega Z [\rho \exp(-F_p / kT)] \quad . \qquad [4]$$

The attachment frequency ω should be taken as the lowest of the two values, one for oxygen and the other for vacancies. The rate-limiting reaction is normally attachment of oxygen, with an exception of low-T range where C_V is strongly diminished due to

vacancy binding into VO_2. The factor Z can be again replaced by a constant value 0.1, and the plate critical radius R_p* – by a constant value 0.3 nm.

Vacancy consumption by voids and oxide plates

The nucleation rate of the dominant clusters (either voids or plates, depending on C_0) increases very fast, upon lowering T, and at some moment the density of the already produced and growing clusters becomes sufficient to diminish the vacancy concentration C_V and thus to suppress further nucleation. In this way the final density of grown-in microdefects is determined. This density is critically dependent on the growth rate of the dominant clusters.

If the growth rate of a void is limited by the vacancy supply by diffusion then the final void density will depend on the cooling rate, $q = -dT/dt$, by a power law $q^{3/2}$ (25). This law was well confirmed experimentally (4). It is important, that in the course of diffusion-limited growth, the oxygen film at the void surface can cover only a fraction of the total surface area (29) because the vacancy flux far exceeds that of oxygen. A large fraction of the void surface is "naked" – not covered by the oxygen film. With a fast two-dimensional nucleation of new vacancy layers (islands) at the naked parts of the facets, the vacancy attachment to a void is diffusion-limited. The vacancies, arriving to the facets, move by surface diffusion and get incorporated into the islands. Similarly, the oxygen atoms arriving to the void surface, move by surface diffusion and get incorporated into the oxygen film. This mode of void growth persists for some time but later the oxygen-covered fraction increases and finally approaches 1. At this moment, a new void can emerge from the initial void (29) which accounts for double voids. During the critical stage of vacancy consumption – when the grown-in void density is determined – the voids are still essentially single-octahedral; a possible morphological change to double voids occurs later. This is a convenient feature to calculate the void density. The diffusion flux of vacancies to a void is equal to $4 \pi R D_V C_V$. Strictly speaking one should use, instead of C_V, the difference between C_V and the near-surface equilibrium vacancy concentration; however, the latter quantity is very small in comparison to C_V. The vacancy capture radius R is generally dependent on the void shape; for a single-octahedral void, R is very close (29) to the effective void radius R_v (a radius of a sphere of the same volume). The total number of vacancies in the void, $(4\pi/3)R_v^3 \rho$, increases due to the incoming diffusion flux, and $R_v(t)$ is described by a simple equation

$$dR_v^2/dt = 2 D_V C_V / \rho \quad . \qquad [5]$$

The growth of plates is assumed to be limited by diffusion of oxygen (and at low T – by diffusion of free vacancies). The oxygen atoms arrive either directly to the plate edge, or first to the plate surface - and then move to the edge by surface diffusion. The fluxes for both cases are not much different; we assume the second mode of collecting oxygen – through the whole surface of a round plate, of a radius R_p. The diffusion flux for this case (sinking to a disc) is known to be $Q_{ox} = 8 R_p D_{ox} C_{ox}$. In this expression, we neglect the near-surface equilibrium concentration of oxygen which is very low in the presence of supersaturated vacancies. The number n of clustered oxygen atoms is $\pi R_p^2 \rho_{ox}$, where ρ_{ox} is the surface density of oxygen in the plate. The plates reside in (001) plane, and we further assume that the VO_2 units replace the Si atoms in a single (001) plane where the Si atom density is $2/a^2$ (a is the lattice constant). The surface density of oxygen in the

plate is twice as large, $\rho_{ox} = 4/a^2 = \rho^{2/3} = 1.36\text{x}10^{15}$ cm^{-2}. The balance equation, $dn/dt = Q_{ox}$, defines the time evolution of the plate radius:

$$dR_p/dt = (4/\pi)\, D_{ox}\, C_{ox} / \rho_{ox} \quad . \qquad [6]$$

The growing plates consume vacancies in proportion to the incoming oxygen flux Q_{ox} (one vacancy per two oxygen atoms).

A void (or a plate) nucleated at some moment t' grows to a radius $R_v(t', t)$ (or $R_p(t', t)$) by the current moment t. These functions are obtained by integration of the kinetic equations [5] and [6]. The current total vacancy concentration C_{Vt} (of V + VO + VO_2) evolves by the loss equation:

$$dC_{Vt}/dt = -\,4\pi D_V C_V \int I_v(t')\, R_v(t', t)\, dt' \; - 4\, D_{ox} C_{ox} \int I_p(t')\, R_p(t', t)\, dt' \quad . \qquad [7]$$

The integration over t' is performed from some reference moment (when both nucleation rates are still negligible) to the current moment t. Due to the loss, C_V changes along the crystal axis, which results in the vacancy redistribution by diffusion. This effect is neglected in Eq.[7]; more sophisticated – and time-consuming - simulations, including axial diffusion, were checked to give almost the same results. The equations [5] – [7] are solved numerically using the expressions [2] and [4] for the nucleation rates. In this way all the relevant quantities are traced in dependence of time – and thus in dependence of the gradually decreasing temperature, with a constant cooling rate $q = -dT/dt$.

The three vacancy species (V, VO and VO_2) co-exist in the equilibrium ratio that is expressed through the current temperature T and the binding temperature T_b (18).

Simulated properties of voids and oxide plates

To perform simulations, the vacancy properties should be defined. The equilibrium concentration C_{Vm} (at T_m) is somewhat uncertain; the presently adopted value is $7\text{x}10^{14}$ cm^{-3}. The formation and migration energy of vacancy are close to 4 eV and 0.4 eV, respectively (30). The vacancy diffusivity D_V (expressed through the reference value at the melting point, D_{Vm}) is the main parameter that controls the void density, and therefore D_{Vm} should be fitted to reproduce a typical measured void density for the standard V-mode – about $1.7\text{x}10^6$ cm^{-3} (4). The fitted value is $D_{Vm} = 1.2\text{x}10^{-4}$ cm^2/s - very close to the value of D_{Vm} obtained from the vacancy profiles installed into wafers by Rapid Thermal Annealing (RTA) (5, 30).

The energy parameter u, of the oxygen-covered void surface, controls the nucleation temperature T_n of voids (the temperature at which $I_v(T)$ is sharply peaked). The experimental value, $T_n = 1100^oC$ (for a standard V-mode), is reproduced at $u = 1.15$ eV. Finally, the edge energy of plates, λ_p, controls the properties of the P-band; the best value was found to be 2.95 eV.

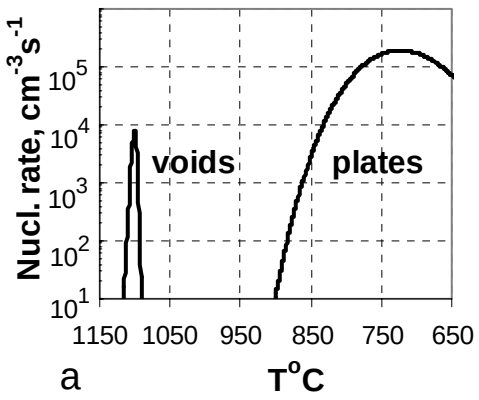
Now, with all the parameters specified, the microdefect properties can be calculated in dependence of the incorporated vacancy concentration C_0, the cooling rate q and the oxygen concentration C_{ox}. In the present study we concentrate on the dependence on C_0. The other two parameters are fixed at representative values for the CZ growth: q = 2 K/min, and $C_{ox} = 7\text{x}10^{17}$ cm^{-3} (this number corresponds to the calibration coefficient $2.45\text{x}10^{17}$ cm^{-2} for the major oxygen IR absorption band).

The computed nucleation curves $I_v(T)$ and $I_p(T)$ are shown in Fig.1 for two representative values of the pre-agglomeration vacancy concentration C_0.

Ex. 1016, Page 977

One value (C_0 = 7x10^{13} cm^{-3}, Fig.1a) corresponds to a standard V-mode when the dominant vacancy clusters are voids. The void nucleation rate I_v is sharply peaked which means that the size distribution of voids is very narrow: almost all the voids are of the same radius, in the present case 58 nm. Some small fraction of vacancies survives the stage of consumption by voids and exists in the form of VO_2; the reason is that a fraction of free vacancies V in the V+VO+VO_2 community is reduced upon lowering T, and vacancy sinking to voids is suppressed. Due to the residual vacancy species, the nucleation rate of plates in Fig.1a increases and has a maximum at 740°C; a broad nucleation curve implies a broad size distribution of plates.

The other value of C_0 (2x10^{13} cm^{-3}, Fig.1b) corresponds to the P-band – when the major vacancy loss is to the plates. A relatively narrow peak in $I_p(T)$ means that the size distribution of plates is also narrow; the average radius is about 75 nm. The voids are still present, but have a small radius, 7.5 nm. The calculated plate density, 1.6x10^8 cm^{-3}, is essentially lower than in the previously simulated case of spherical precipitates (10); now it is close to the experimental precipitate density in the P-band. The reason for a lower density is that the plates are much more efficient collectors of oxygen (and thus of vacancies) in comparison to spherical precipitates: the plate radius R_p increases quickly with time, in proportion to t while the radius of a spherical particle increases in proportion to $t^{1/2}$ and remains much smaller than R_p. A relatively low density of growing plates is thus sufficient for efficient vacancy consumption - and for suppressing further plate nucleation.



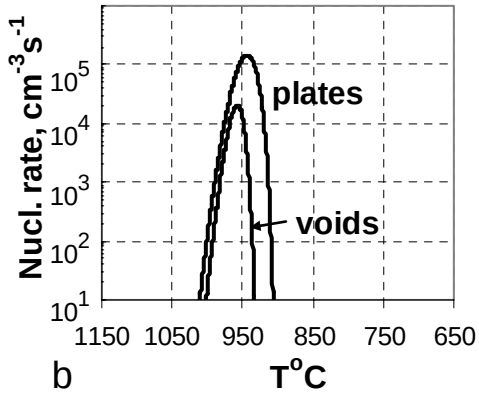


Fig.1 Nucleation rate of voids and plates in dependence of decreasing T; (a) at C_0 = 7x10^{13} cm^{-3}, (b) at C_0 = 2x10^{13} cm^{-3}.

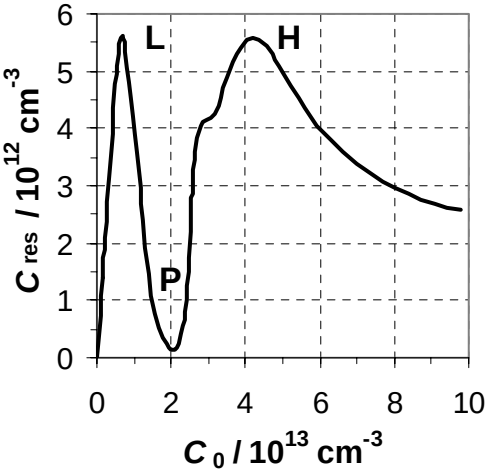


Fig.2 Grown-in (residual) vacancy concentration, in dependence of the initial vacancy concentration that existed before the onset of clustering.

The grown-in value of the residual vacancy concentration C_{res} is shown in Fig.2 in dependence of the initial vacancy concentration C_0. The P-band - a region of relatively large plates - corresponds to the minimum in $C_{res}(C_0)$. The minimum is surrounded by two peaks marked H and L (H stands for Higher C_0, L stands for Lower C_0). The corresponding two spatial bands, H-band and L-band, surround the P-band. Within the H- and L- bands, numerous small plates are present, with a broad size

distribution. The distribution function is shown in Fig.3 for the L-band (C_0 = 7x10^{12} cm^{-3}). Instead of the differential function, dN_p/dR_p (that shows the number of defects per unit size interval) a more friendly cumulative function $N_p(R_p)$ is used; it shows the density of defects with a size larger than R_p. When a silicon sample is annealed, only the plates larger than some critical size R_p* will grow; the smaller ones will dissolve. From the plot of Fig.3, one can directly see the density of plates that will grow, and finally develop into large precipitates (amorphous platelets). The critical radius R_p* is obtained from the expression for the plate free energy $F(m)$ in dependence of the number of VO_2 units, for a particular case of equilibrium vacancy concentration (μ_V = 0):

$$R_p^* = [\lambda_p / (2\,\mu_{ox} - f_s)] / (4\,\pi\,\rho_{ox})^{1/2} \quad . \qquad [8]$$

With the currently adopted parameters of plates (f_s = 0.4 eV, λ_p = 2.95 eV, ρ_{ox} = 1.36x10^{15} cm^{-2}) and C_{ox} = 7x10^{17} cm^{-3}, the critical radius is 2.3 nm for the conventional precipitation anneal at 1000°C. The critical radius is very sensitive to T. At 1025°C, R_p* = 4.2 nm (hence a major part of the grown-in plates will dissolve). At T > 1050°C the solubility C_{pe} becomes higher than C_{ox}, and all the plates will dissolve. Why then the oxide particles of the P-band survive the oxidation anneal at essentially higher T ? The answer is that the plates in the P-band are so large that they transform into the stable form, amorphous platelets, already during crystal cooling.

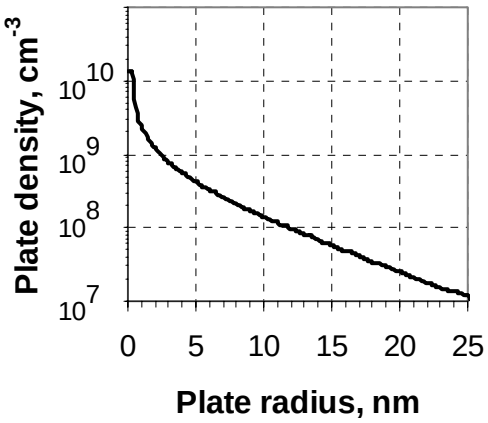


Fig.3 Density of plates with a radius larger than the specified value. This distribution is for the L-band.

At 800°C (the temperature used to stabilize the precipitate nuclei), R_p* = 0.46 nm (m = 9). According to Fig.3, almost all grown-in plates evolve to a larger size. This means that after prolonged anneal at around 800°C the local precipitate density in the H- and L-bands is much higher than in all other parts. Two rings of heavy precipitation will be seen in a wafer; one inside the OSF ring (the H-band), and the other outside (the L-band).

The computed microdefect density in dependence of C_0 is presented in Fig.4. The plate density is shown for relatively large plates – those with a radius exceeding 3 nm (and thus capable to survive 1000°C anneal). Within the P-band, this density is identical to the total plate density, but outside the P-band, it is only a fraction of the total density.

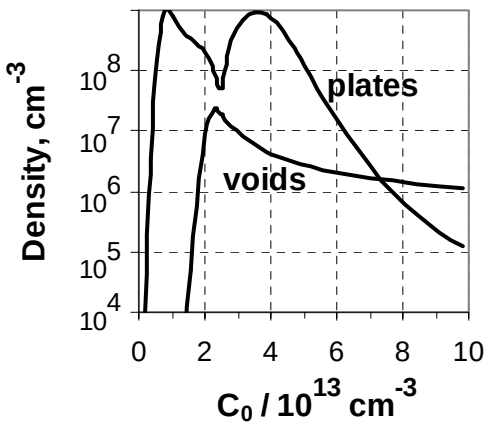


Fig.4 Void and plate density in dependence of pre-agglomeration vacancy concentration C_0. Only the plates with R_p > 3 nm are included.

Double-octahedral voids induced by oxygen adsorbed film

The double-octahedral voids are found (19, 20) when the oxygen concentration C_{ox} is high and/or the cooling rate q is low. At low C_{ox} and high q, the voids are single-octahedral. This observation alone suggests that the original void shape is single-octahedral, and that the morphological change to double-octahedral structure is caused by oxygen. It was already mentioned that the void facets, in the initial stage of void growth, are partially covered by the oxygen film. The total void surface area S_v increases according to Eq.[5] – in proportion to the time t. The oxygen-covered area S_{ox} increases in proportion to the oxygen flux (which is in turn proportional to void radius R_v) and hence S_{ox} increases in proportion to $t^{3/2}$. At longer time, S_{ox} will increase faster than S_v, and the film will finally cover the total void surface. Instead of time, it is more convenient to use the consumed vacancy fraction $\delta = 1 - C_{Vt}/C_0$. The two areas, S_v and S_{ox}, are simple functions of δ (29); S_v is just proportional to $\delta^{2/3}$, and $S_{ox}(\delta)$ is dependent on some single parameter P combined of the three growth parameters: P is proportional to $C_0^{1/2}\, q^{1/2} / (C_{ox} - C_e^*)$ where C_e^* is the oxygen concentration in equilibrium with the adsorbed oxygen film; C_e^* somewhat exceeds the regular solubility C_e. The parameter P is large for a high q and/or low C_{ox} – and small otherwise. The evolution of S_v and S_{ox}, represented by the dependence on δ, is plotted in Fig.5, for $P = 1$. The moment of complete coverage corresponds to the intersection of the two curves. For larger P the complete coverage occurs at δ close to 1; in other words, the remaining vacancy concentration is low. For smaller P it occurs at small δ; the remaining vacancy concentration is high and close to the initial value C_0. The experimental rule for the void type (single or double) can be expressed in terms of the value of P (29): single if $P > 2$ and double if $P < 2$. In other words, the fate of the initial void – either to remain single or to become double – is decided by the remaining vacancy concentration at the moment of complete coverage of the void surface with oxygen film.

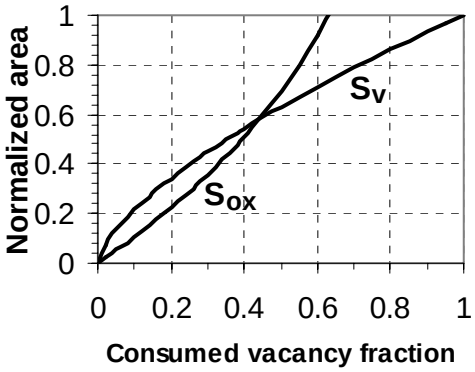


Fig.5 Evolution of the total void area S_v and the area S_{ox} covered by oxygen film. Both S_v and S_{ox} are normalized by the final void area.

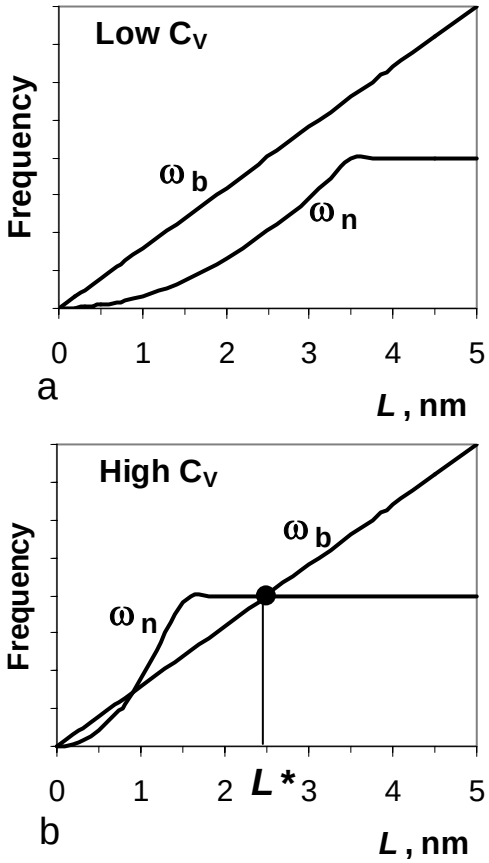


Fig.6 The frequencies ω_n (for nucleating new vacancy layers) and ω_b (for overcoming the boundary between the naked and oxygen-covered parts), in dependence of the size L of remaining naked part.

The key importance of the remaining vacancy concentration C_V can be understood in the following way. The new vacancy layer (island), nucleated at the naked part of a facet, propagates laterally by incorporating the vacancies that arrive by surface diffusion. This process can be described as the lateral movement of the edge of the island (called a step). When the step arrives at the boundary between the naked part and the oxygen-covered part, it can temporarily stop – if the boundary is an obstacle for the further lateral movement through the oxygen-covered part. The boundary is overcome by 1D nucleation – through a local protrusion of a small step segment into oxygen-covered part. The probability to overcome the boundary per unit time, ω_b, is then proportional to the step length L. The size of the naked part is in the order of L. The islands are produced one by one, by 2D nucleation, at some frequency ω_n proportional to L^2 (the area of nucleation) if L is sufficiently small. At larger L this frequency is determined by diffusion supply of vacancies to the void, and thus becomes independent of L.

A schematic dependence of the characteristic frequencies, ω_n and ω_b, on the size L of the remaining naked part, is shown in Fig.6. The frequency ω_n is very small at low C_V (Fig.6a), and remains much lower than ω_b. It means that a step, that has arrived to the boundary, escapes into the oxygen-covered part long before the next step arrives. Under these conditions the steps move independent one of the other, the oxygen-covered part propagates to the whole facet - and a single-octahedral void, completely covered with oxygen film, emerges. At higher C_V, however, the nucleation frequency is much higher, and the $\omega_n(L)$ curve can intersect the other curve, $\omega_b(L)$, as shown in Fig.6b. When the size L of still remaining naked spot becomes smaller than some characteristic value L^* (see Fig.6b), the new step arrive to the boundary before the previous step has escaped. The steps are thus piled up at the boundary, and a new void, initially of a size L^*, protrudes from the old void as shown schematically in Fig.7. The two voids are connected by a small hole (7-9); this is possible only if the internal facets of the double structure – those adjacent to the hole (and marked *i* in Fig.7) do not move or move at a low velocity. It can be concluded that

(1) the facets, completely covered by oxygen film, are immobile (or move only slowly),

(2) the internal facets of the new void are completely covered by oxygen film, as shown in Fig.7, and belong to the slowly moving type.

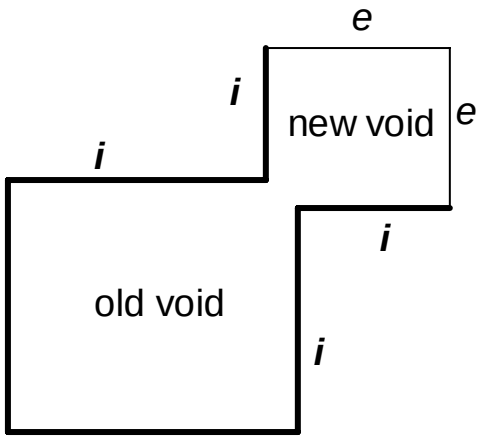


Fig.7 A new void emerged from the last naked spot at the corner of the old void, by piling-up the steps. The facets covered by oxygen film are shown by thick lines, and those non-covered – by thin lines.

The second condition is maintained if the mode of incorporation of oxygen and vacancies by the new void is as follows:

(a) the oxide film easily spreads from the old void facets to the adjacent facets of the new void (Fig.7), and the film continues to incorporate, by surface diffusion, the oxygen atoms arriving to the old void,

(b) the new void collects vacancies only through its own surface, and does not receive them from the old void (in other words, the adsorbed vacancies can not penetrate through the boundary between the two voids).

Under these conditions, the surface area S_v of the new void increases according to Eq.[5] – at the same rate as that for the old void before doubling. Also the oxygen supply

to the new void is the same as that to the old void before doubling. The oxygen-covered area S_{ox} was increasing faster than S_v (Fig.5) which means that the oxygen flux is sufficient to keep the internal facets of the new void completely covered. The structure shown in Fig.7 also implies that the oxygen film can not spread from the internal facets of the new void to the external facets (marked *e* in Fig.7); otherwise also the external facets would be covered by the oxygen film. In short, in the double-void structure, only the external facets of the new void move at a fast (diffusion-limited) velocity; the movement of all the other facets is blocked by the oxygen film.

Platelet-shaped voids induced by nitrogen adsorbed film

For moderate doping level with nitrogen, the voids in CZ crystals are platelets (the thickness is about 11 nm, the diameter is about 80 nm) (21). The void density is increased significantly by nitrogen doping while the void volume is reduced. The nitrogen effect on the void density was initially ascribed (24) to vacancy trapping by nitrogen; the major traps should be single interstitial atoms N_1 while the dominant species, interstitial dimers N_2, have too small binding energy to induce significant trapping in the temperature range of void formation in CZ crystals. It has turned out, however, that trapping of vacancies by N_1, although favored thermodynamically, proceeds at a negligible rate (31). The only mechanism that can account for the nitrogen effect on voids is then formation of adsorbed nitrogen film at the void surface. This effect is quite similar to the previously discussed case of oxygen film at the void facets. The nitrogen diffusion flux to a void (through N_1 mobile species) is much greater than that of oxygen (31), and this is the main reason for the strong effects due to nitrogen. We assume that nitrogen film is not present initially (contrary to the oxygen film) but nucleated at the naked parts of void facets during the initial stage of diffusion-limited growth.

The platelet morphology is accounted for within the following scenario. The adsorbed nitrogen atoms penetrate from a facet to an adjacent facet, but not quite easily. After the nitrogen island is nucleated, randomly at one of the facets, it collects adsorbed nitrogen both from its own facet and from the adjacent facets, and the surface nitrogen concentration at these facets is significantly reduced - which prevents nucleation of nitrogen film at these facets. The facet opposite to that containing an island of nitrogen film, is less depleted with nitrogen, and nucleation of another island of a film is possible. The resulting structure, with two islands of nitrogen film at two opposite facets, is shown schematically in Fig.8a.

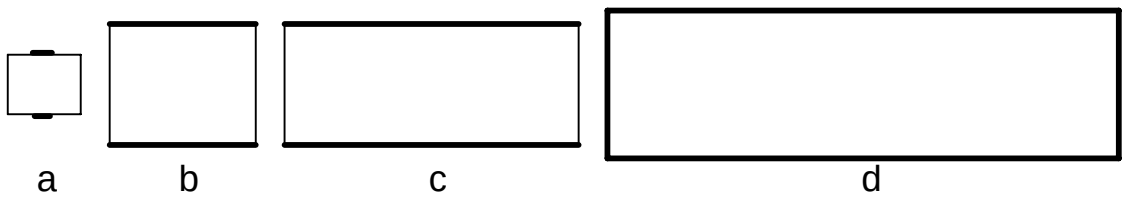


Fig.8 Evolution of void shape in nitrogen-doped CZ crystal. (a) cubic void with two islands of nitrogen film at two opposite facets, (b) cubic void at the moment of complete coverage of the two facets by the film, (c) platelet void with two blocked basal facets, (d) platelet void with all the facets blocked.

For the sake of simplicity, cubic shape of a void (with 6 facets) is considered. The void size is defined as the distance *R* from the void centre to a facet. The number of

Ex. 1016, Page 982

consumed vacancies, $8R^3\rho$, is proportional to $D_V C_V$. The number of nitrogen atoms in the nitrogen film is equal to $\rho_n S_n$, where ρ_n is the surface density of nitrogen atoms in the film, and S_n is the total area of the two nitrogen islands. This number is proportional to $D_1 C_1$ – the product of the diffusivity and concentration for N_1 species. Hence

$$S_n = (8 R^3 \rho / \rho_n) (D_1 C_1 / D_V C_V) \quad . \qquad [11]$$

The complete coverage of the two opposite facets (Fig.8b) is achieved when the area $S_n/2$ of the film, at each facet, becomes equal to the facet area, $4 R^2$. By Eq.(11), it occurs at some definite void size R:

$$R = \rho_n D_V C_V / (\rho D_1 C_1) \qquad . \qquad [12]$$

Starting from this moment, the growth of two nitrogen-covered facets is blocked, and only four other facets continue to move. The cubic shape will then evolve into a rectangular platelet, with two basal immobile facets and four moving lateral facets (Fig.8c). The thickness of the platelet, w, is twice as large as the radius R defined by Eq.[12]. A typical value of $D_V C_V$ is $4x10^9$ $cm^{-1}s^{-1}$, the product $D_1 C_1$ is expressed through the total nitrogen concentration C_n using a well-defined dissociation parameter of nitrogen diffusion (31); the estimated product is $1.6x10^8$ $cm^{-1}s^{-1}$ for $C_n = 10^{15}$ cm^{-3} and $T = 1100^oC$. The nitrogen density in the film is roughly taken as the surface Si density, $8x10^{14}$ cm^{-2}. Then the calculated thickness $w = 2R = 8$ nm, close to the observed value.

At the initial stage of still cubic void (Fig.8a), the facet area evolves in proportion to the time t. The area of the nitrogen film increases in proportion to $t^{3/2}$ (similar to the oxygen-covered area in the previous section). At the moment of complete coverage (Fig.8b) the lateral velocity of the film is larger than the facet velocity by the ratio of the two exponents, 3/2. After the growth of nitrogen-covered facets is blocked, the same diffusion-limited vacancy flux – previously uniformly accommodated at 6 facets – is passed to only 4 still moving facets. Therefore, the velocity of these 4 facets is increased, in a jump, by a factor of 6/4, and becomes identical to the lateral velocity of the nitrogen film. The structure with completely covered two basal facets will be thus reproduced in the process of lateral growth (Fig.8c) if the basal facets were immobile. Actually, these facets can move in the normal direction at some low velocity. The platelet thickness w will then gradually increase, and the lateral facet velocity will be accordingly reduced. As a result, the film spreads from the basal facets into the lateral facets, and finally also the lateral facets become completely covered and blocked (Fig.8d). A complicated problem of the final void radius R_v is to be addressed separately. It can be expected that R_v is proportional to the initial thickness w. According to Eq.(12), $w = 2R$ is inversely proportional to C_1, and thus to the square root of the total nitrogen concentration C_n. The volume Ω of just produced platelet void (in the order of $4 R_v^2 w$) will depend on C_n as $1/C_n^{3/2}$. The nucleated voids reach the state of blocked platelets (Fig.8d) within a short time, due to a fast lateral growth of a platelet. Qualitatively, it can be said that the void nucleation results in instantaneous production of platelet voids, of a specified volume Ω. The observation of voids in quenched (32) and halted (33) nitrogen-doped crystals suggests that the voids become detectable only after the crystal is cooled from the nucleation temperature (1100^oC) to some lower temperature (like 1060^oC). The just produced platelets are thus too thin to be detected, but by a slow propagation of the two basal facets, the voids become gradually thicker, and finally reach the size of detection.

Summary

A better reproduction of the grown-in void and oxide particle properties (in comparison to the previous calculations) is achieved using two important modification of the model:
(1) The oxide precipitates are thin plates rather than spherical particles. Due to a faster lateral growth of plates, the plate density, sufficient to reduce the vacancy concentration and to suppress further nucleation, is lower than in the case of spherical particles. The calculated density of grown-in precipitates, in the P-band, is now about 10^8 cm^{-3} which is close to the experimental density.
(2) A higher value is assumed for the binding temperature (T_b) of vacancies by oxygen into VO and VO_2 species: 1100°C instead of the previous value of 1050°C. The void density is controlled by the effective vacancy diffusivity D_V* that depends on the fraction of V in the V+VO+VO_2 community. At the previously assumed lower T_b, the effective (apparent) diffusivity D_V*, at a typical void nucleation temperature of 1100°C, was close to the true diffusivity D_V. The deduced D_V was essentially lower than the value based on the RTA-induced vacancy profiles and extrapolated to 1100°C (30). The now adopted higher T_b is close to the formation temperature of voids, and D_V becomes essentially larger then the apparent value D_V* - and now consistent with the RTA-based vacancy diffusivity.

A complicated void shape – different from a simple octahedron – is accounted for by adsorbed impurity film at the void facets. Oxygen film is present initially - at the facets of nucleated voids; in the course of void growth, the film area is first reduced to a small fraction of the total area but then again expands to the total void area. At this moment – if the oxygen content is sufficiently high - a morphological change to a double-octahedral structure occurs. Nitrogen film (in nitrogen-doped CZ crystals) is not present initially; it is nucleated later, only at two opposite facets, and blocks the growth of these facets. After that, the void propagates only laterally and becomes a platelet.

References

1. V. V. Voronkov, *J. Crystal Growth*, **59**, 625 (1982).
2. W. von Ammon, E.Dornberger, H. Oelkrug, H. Weidner, *J. Crystal Growth*, **151**, 273 (1995).
3. M. Hourai, E. Kajita, T. Nagashima, H. Fujiwara, S. Umeno, S. Sadamitsu, S. Miki and T. Shigematsu, *Materials Science Forum*, **196-201**, 1713 (1995).
4. T. Saishoji, K. Nakamura, T. Yokoyama, F. Ishikawa and J. Tomioka, in *High Purity Silicon V*, C. L. Claeys, P. Rai-Choudhuri, M. Watanabe, P. Stallhofer and H. J. Dawson, Editors, PV 98-13, p. 28, The Electrochemical Society Proceedings Series, Pennington, NJ (1998).
5. V. V. Voronkov and R. Falster, in *High Purity Silicon 10*, C. L. Claeys, R. Falster, M. Watanabe and P. Stallhofer, Editors, PV 16, no. 6, p. 15, The Electrochemical Society Proceedings Series, Pennington, NJ (2008).
6. V. V. Voronkov and R. Falster, *J. Appl. Phys.*, **86**, 5975 (1999).
7. T. Ueki, M. Itsumi and T. Takeda, *Jpn. J. Appl. Phys.*, **37**, 1667 (1998).
8. M. Nishimura, S. Yoshino, H. Motoura, S. Shimura, T. Mchedlidze and T. Hikone, *J. Electrochem. Soc.*, **143**, L243 (1996).
9. M. Kato, T. Yoshida, Y. Ikeda and Y. Kitagawara, *Jpn. J. Appl. Phys.*, **35**, 5597 (1996).
10. V. V. Voronkov and R. Falster, *J. Crystal Growth*, **204**, 462 (1999).

11. M. Hasebe, Y. Takeoka, S. Shinoyama and S. Naito, in *Defect Control in Semiconductors 1*, S. Sumino, Editor, p. 157, Elsevier, Amsterdam (1990).
12. H. Shimizu, C. Munacata, N. Honma, S. Aoki and Y. Kosaka, *Jpn. J. Appl. Phys.*, **31**, 1817 (1992).
13. K. Marsden, S. Sadamitsu, T. Yamamoto and T. Shigematsu, *Jpn. J. Appl. Phys.*, **34**, 2974 (1995).
14. V. V. Voronkov, *Materials Science & Engineering*, **B73**, 69 (2000).
15. M. S. Kulkarni, *J. Crystal Growth*, **303**, 438 (2007).
16. W. Bergholz, M. J. Binns, G. R. Booker, J. C. Hutchison, S. H. Kinder, S. Messolaras, R. C. Newman, R. J. Stewart and J. G. Wilkes, *Philos. Mag.*, **B59**, 499 (1989).
17. R. Falster, V. V. Voronkov, V. Y. Reznik and M. G. Milvidski, *Solid State Phenomena*, **108-109**, 97 (2005).
18. V. V. Voronkov and R. Falster, *J. Electrochem. Soc.*, **149**, G167 (2002).
19. S. Umeno, Y. Yanase, M. Hourai, M. Sano, Y. Shida and H. Tsuya, *Jpn. J. Appl. Phys.*, **38**, 5725 (1999).
20. F. Ishikawa, S. Sadonara, T. Saishoji, K. Nakamura and J. Tomioka, in *High Purity Silicon VI*, C. L. Claeys, P. Rai-Choudhuri, M. Watanabe, P. Stallhofer and H. J. Dawson, Editors, PV 2000-17, p. 86, The Electrochemical Society Proceedings Series, Pennington, NJ (2000).
21. J. Takahashi, K. Nakai, K. Kawakami, Y. Inoue, H. Yokota, A. Tachikawa, A. Ikari and W. Ohashi, *Jpn. J. Appl. Phys.*, **42**, 363 (2003).
22. S. Umeno, S. Sadamitsu, H. Murakami, M. Hourai, S. Sumita and T. Shigematsu, *Jpn. J. Appl. Phys.*, **32**, L699 (1993).
23. K. Nakamura, T. Saishoji and J. Tomioka, in *High Purity Silicon VIII*, C. L. Claeys, M. Watanabe, R. Falster and P. Stallhofer, Editors, PV 2004-05, p. 237, The Electrochemical Society Proceedings Series, Pennington, NJ (2004).
24. V. V. Voronkov and R. Falster, *J. Crystal Growth*, **273**, 412 (2005).
25. V. V. Voronkov and R. Falster, *J. Crystal Growth*, **194**, 76 (1998).
26. G. Kissinger and J. Dabrowski, *J. Electrochem. Soc.*, **155**, H448 (2008).
27. J. Mikkelsen, Mater. Res. Soc. Symp. Proc., 59, 19 (1986).
28. R. C. Newman, M. J. Binns, W. P. Brown, F. M. Livingston, S. Messolaras, R. J. Stewart and J. G. Wilkes, Physica B+C, 116, 264 (1983).
29. V. V. Voronkov and R. Falster, *J. Crystal Growth*, **226**, 192 (2001).
30. V. V. Voronkov and R. Falster, *Materials Science & Engineering*, **B134**, 227 (2006).
31. V. V. Voronkov and R. Falster, *Solid State Phenomena*, **131-133**, 219 (2007).
32. K. Nakamura, T. Saishoji, S. Togawa and J. Tomioka, in *Proc. of the Kazusa Academia Park Forum on the Science and Technology of Silicon Materials*, the Japan Technical Information Service, p.109 (2001).
33. K. Nakai, Y. Inoue, H. Yokota, A. Ikari, J. Takahashi and W. Ohashi, in The 3-d International Symposium on Advanced Science and Technology of Silicon Materials, The Japan Society for Promotion of Science, p.88 (2000).

ECS Transactions, 18 (1) 959-967 (2009)
10.1149/1.3096559 © The Electrochemical Society

Atomistic-to-Continuum Modeling of Defect-Related Phenomena in Silicon Crystals

T. Sinno and S. S. Kapur

Department of Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Continuum process modeling of point defect and impurity aggregation during silicon crystal growth and wafer annealing has led to significant contributions toward understanding and improvement of industrial processes. Key inputs to these models are thermophysical properties of point defects and their clusters, which may be strong functions of temperature and cluster size. In this paper, a theoretical framework is presented for probing the high-thermodynamic properties of vacancy and self-interstitial clusters in crystalline silicon at elevated temperature. In particular, configurational and vibrational entropy are shown to be significant in both types of defect clusters. In both cases, configurational entropy leads to the existence of a wide distribution of possible cluster configurations that collectively lowers the free energy of formation relative to the energetic ground state configuration. Moreover, certain self-interstitial cluster sizes are additionally stabilized by configurations that possess anomalously high vibrational entropy.

Introduction

The quality of polished silicon wafers in the context of microelectronic device fabrication is strongly affected by the density and size distribution of native point defect aggregates. These clusters result from the aggregation of point defects that become supersaturated as a result of the cooling encountered during crystal growth (1). Moreover, some device fabrication steps, such as ion implantation, also often lead to highly non-equilibrium conditions that produce high point defect supersaturation and clustering (2). As a result, numerous studies within the literature have focused on modeling and simulation of point defect aggregation during crystal growth and wafer processing (3,4,5).

It is now well established that under low-dopant conditions, native point defect clustering into large aggregates generally proceeds by homogenous nucleation and growth. Within this framework, continuum-scale process models, with appropriate values for point defect and cluster thermodynamic and transport properties, are able to reproduce the general features of native defect cluster formation. On the other hand, continuum models are extremely sensitive to the values of the thermophysical parameters (e.g. diffusion coefficients and equilibrium concentrations), and it has proven to be extremely difficult to independently establish these values to a degree that provides quantitative prediction of experimental measurements of defect sizes and densities. This is particularly true in the case of small defect clusters, whose properties are difficult to isolate and measure experimentally.

Ex. 1016, Page 987

In this paper, we consider the properties of small defect clusters using atomistic scale simulations and show that even native point defect clusters can exhibit complex thermodynamic behavior, which must be captured accurately if quantitative predictions of defect size distribution evolution are to be made at the continuum level. In the following section, a general framework is presented for analyzing the properties of defect clusters at finite temperature. The essence of the approach is to probe the potential energy landscape (PEL) of a given defect cluster by employing lengthy molecular dynamics simulations (6,7). This information is then used to compute the total (classical) free energy of the defect as a function of temperature.

Potential Energy Landscapes of Defects in Crystals Using Molecular Dynamics

In general, the free energy of formation for a crystal defect is determined by its enthalpy and two potential sources of entropy, namely vibrational and configurational. The former entropic contribution can be readily computed for a given structure using a number of techniques (e.g. quasi-harmonic analysis (QHA) (8)) and is usually included in the formation free energy of solid-state defects. However, the latter is significantly more challenging to access because it arises from the presence of a *collection* of configurations, all of which must be identified before the total configurational entropy can be estimated. Simple defect analyses (e.g. for vacancy clusters) have attempted to estimate defect configurational entropy on the basis of lattice models, but these are only valid if all important configurations can be defined on a rigid lattice. In fact, at elevated temperature, both vacancy and interstitial clusters in silicon can assume numerous configurations that are not aligned with the underlying silicon lattice – such configurations are impossible to capture in a lattice-based analysis.

In the following analysis, the need for analytically estimating configurational degrees of freedom is bypassed. A defect cluster consisting of n defects is placed into a cubic simulation box containing N silicon lattice atoms. The system is evolved with molecular dynamics (MD) in the constant NVT ensemble, in which the temperature and volume of the system are fixed. In this work, the Environment-Dependent Interatomic Potential (EDIP) for silicon (9) is used for all calculations, and time integration is performed using the fifth-order Gear predictor-corrector method. During the MD runs, the system coordinates are periodically quenched to the local minimum potential energy using a conjugate-gradient minimization approach. The formation energy of the quenched defect configuration, E^f, is calculated as $E^f \equiv E_d - [(N \pm n)/N]E_p$, where E_d is the energy of defect-containing system and E_p is the energy of a reference perfect crystal. The formation energies are collected into bins of width 0.1 eV to form a probability distribution function (PDF).

The PDF, $p(E^f)$, is directly connected to the total formation free energy of the cluster by the relationship $G = -k_B T \ln Z$, in which the configurational partition function is given by (7)

$$Z = \frac{1}{\Lambda^{3N}} \int p(E^f)\, d(E^f) \equiv \frac{1}{\Lambda^{3N}} \int g(E^f) \exp\left(-\beta E^f\right) d(E^f), \qquad [1]$$

where the de Broglie wavelength is defined as $\Lambda = (h^2 / 2\pi m k_B T)^{1/2}$ and $g(E^f)$ is the density-of-states (DOS) for the cluster formation energy. The DOS is a temperature independent quantity that represents the number of states (i.e. mechanically stable configurations) with a given formation energy, and can be extracted directly from the probability distribution function. In principle, the integral in eq. [1] is taken over all possible values of the formation energy but in practice only a narrow range of formation energies are dominant. Note that implicit in eq. [1] is the assumption that the total phase space volume is well approximated by an integral over the local minima in the potential energy landscape. This important assumption is justified for solids, glasses, and supercooled liquids (10).

Vacancy Clusters

Probability distribution functions for several vacancy clusters are shown in Figure 1. In each case, the PDF is a broad function that peaks at a formation energy value that is substantially higher than the ground state value. For example, in the case of the 6-vacancy cluster the energetic ground-state is the well known Hexagonal Ring configuration (shown in panel (a) of Figure 1) with formation energy of 9.8 eV. However, at 1600K this structure is almost never observed during MD simulation. Instead, the most likely structures are those with formation energies of about 11 eV; an example configuration with formation energy of approximately 11.8 eV is shown in panel (b) of Figure 1. These structures tend to involve substantial off-lattice relaxations and therefore cannot be easily represented within a lattice model. Also, the large number of displaced atoms surrounding these clusters implies that their capture volume during growth is larger than that of the ground state configuration (11).

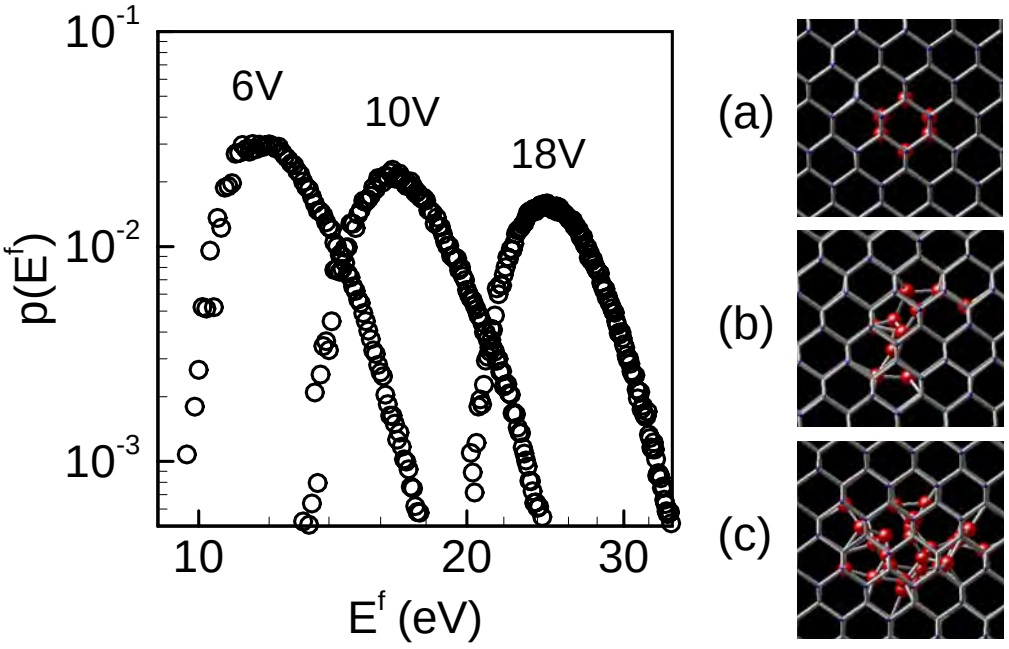


Figure 1. Probability distribution functions for the formation energies of several vacancy clusters at 1600 K. Right panels: Sample configurations for the 6-vacancy cluster at formation energies of (a) 9.8 eV, (b) 11.8 eV, and (c) 16.8 eV, showing increasing disorder with formation energy. Large spheres denote atoms that are displaced by more than 10% of a bond length from their equilibrium positions.

The broad distributions shown in Figure 1 arise from the presence of an extremely large number of mechanically stable cluster configurations, each of which correspond to a local minimum in the potential energy landscape. In fact, the density-of-states increases exponentially with formation energy because at higher formation energies, the larger number of disturbed lattice atoms can lead to a greater variety of mechanically stable configurations. Moreover, as will be shown later, each of these higher energy configurations are further stabilized by vibrational entropy, which also increases as the defect becomes increasingly disordered. The combination of these two entropic features implies that the total free energy of a vacancy cluster is not fully determined by a single structure but rather the entire *distribution* of structures (see eq. [1]). Finally, we note that the presence of entropic effects produces strong temperature dependence in the cluster formation free energy.

The formation free energies for small vacancy clusters based on the PEL analysis are shown in Figure 2 and are compared to those obtained on the basis of the ground state structures alone (upper surface). Note that all free energies are expressed in terms of an effective surface free energy, defined as $\sigma(n,T) = G(n,T) / A_{sph}$, where A_{sph} is the surface area of a sphere with volume equivalent to the sum of the volumes of the vacancies in the cluster. At low temperature, the two estimates converge because entropic contributions are small; in fact, this agreement also serves to validate the accuracy of using eq. [1] to compute defect formation free energies. In the low temperature limit, the cluster free energy is dominated by the energetic ground state and all other possible (higher energy) configurations become unimportant.

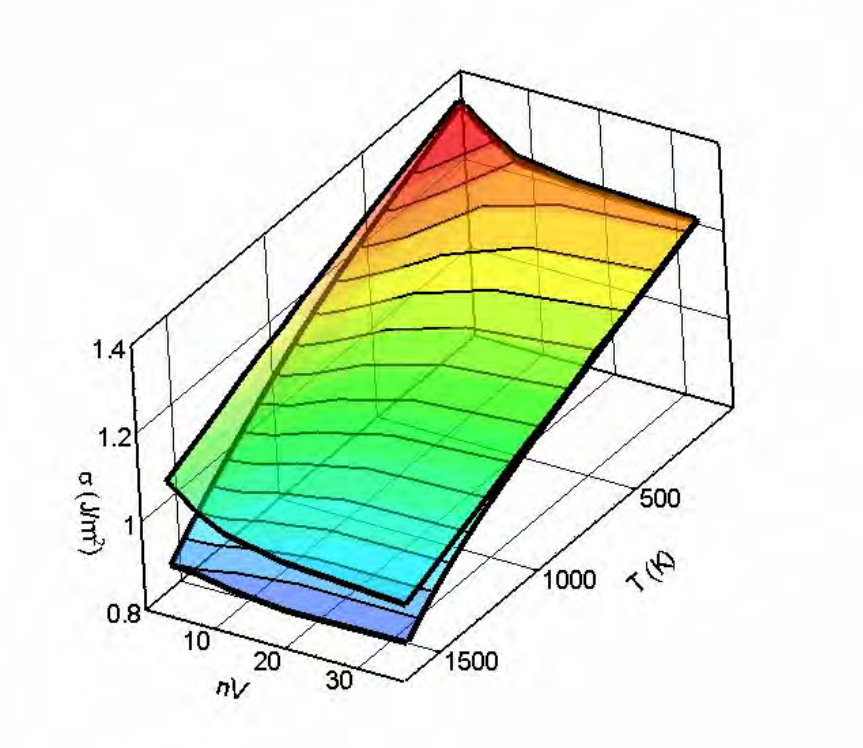


Figure 2. Temperature and size dependence of effective surface free energy for vacancy clusters. Lower surface (shade+lines): PDF results including vibrational and configurational entropy, Upper surface (shade only): HRC calculations with vibrational entropy only.

On the other hand, at elevated temperatures where vacancy clustering is initiated during silicon crystal growth (which is typically around 1100 °C during Czochralski crystal growth (3)), the surface free energies computed with the PEL approach are significantly lower, reflecting the contribution of configurational entropy. Note that both surfaces in Figure 3 include the contribution of vibrational entropy. A qualitatively different aspect of the two surfaces in Figure 3 is the lack of substantial size dependence in the PEL results; while the ground state surface shows a significant rise in free energy per unit area for very small clusters at all temperatures, the PEL result becomes almost flat in the size dimension near the melting temperature. In general, very small defect clusters are known to exhibit energies that deviate from the large cluster limit (usually higher and not necessarily monotonically). On the other hand, the effect of configurational entropy is more significant for small clusters than larger ones (more easily accessible states in the PEL), which counteracts the additional enthalpy resulting in an approximately constant surface energy as a function of size at very high temperature.

With the PEL approach, the effective surface free energy decreases to an approximately constant (in size) value of about 0.8 J/m^2 at the melting temperature, a value that, when applied within continuum process models, leads to improved agreement with experimental measurements of void size distributions in Czochralski-grown silicon crystals (11). Moreover, our results also explain why previous (cryogenic) measurements of the (111) surface free energy of silicon led to much higher values of about 1.2-1.3 J/m^2 (12,13). In this limit, all entropic effects are negligible and large vacancy clusters in their energetic ground state are dominated by the enthalpy of the (111) surface (because of their octahedral morphology), which is predicted by our calculations to be 1.24 J/m^2.

Self-Interstitial Clusters

The overall PEL picture is qualitatively similar for small self-interstitial clusters as shown in Figure 3. These probability distribution functions are evaluated at 1100K, which is within the range of temperatures commonly employed for post-implantation annealing. During this annealing, supersaturated interstitials created by the implant rapidly form clusters, whose distribution evolves by a complex combination of ripening and dissolution as the supersaturation is released at the wafer surface (14). Once again, quantitative modeling of this process requires that the thermodynamic properties of interstitial clusters be well characterized.

The PDFs for the 3- and 5-interstitial clusters are unremarkable in that they, as in the vacancy case, exhibit broad distributions of mechanically stable configurations that are peaked at formation energies substantially higher than the minimum value. This generic trend arises because the increasing density of local minima with formation energy is balanced by the decaying Boltzmann factor in eq. [1], leading to the usual broad peak at an intermediate formation energy. On the other hand, the 4-interstitial cluster PDF exhibits a very large anomalous peak at 8.8 eV. The energy bin centered at 8.8 eV is entirely dominated by the well-known structure shown in panel (a) of Figure 2. Note that the dominance of this structure, which is henceforth referred to as the Humble/Arai configuration (15,16), is not dictated by energetic stability – neighboring structures with similar formation energies are far less likely to be observed during the MD simulations. Moreover, the dominance of this configuration is persistent as the temperature of the

Ex. 1016, Page 991

system is increased to 1300K and beyond. These two observations suggest that the origin of the special stability of this configuration is entropic in nature.

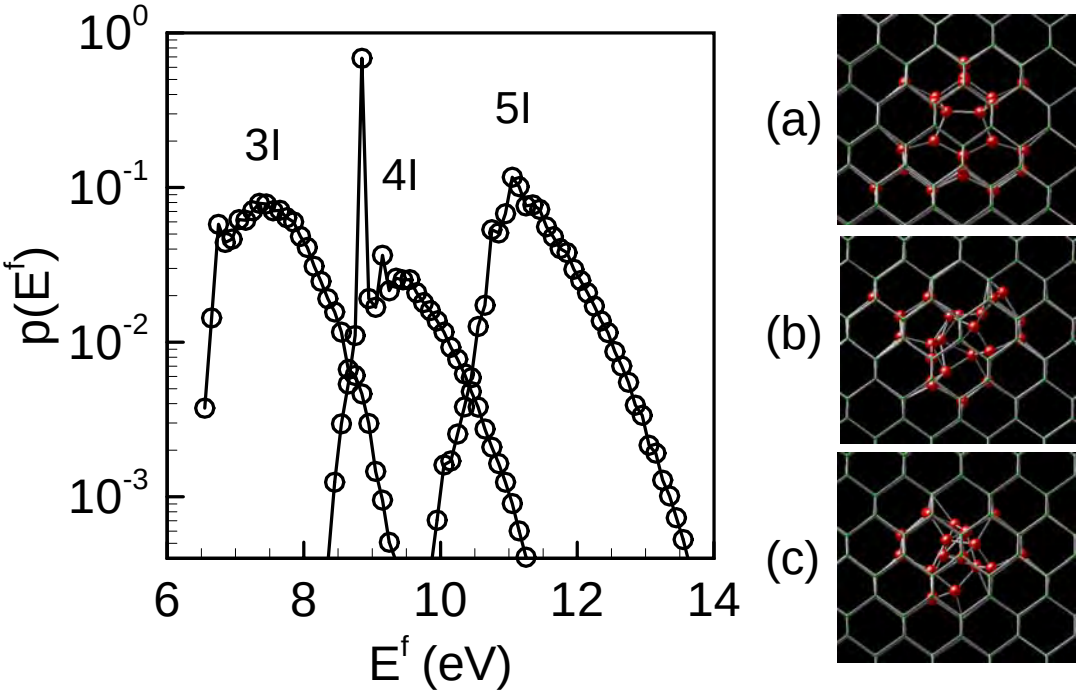


Figure 3. Probability distribution functions for the formation energies of several self-interstitial clusters at 1100 K. Right panels: Sample configurations for the 4-interstitial cluster at formation energies of (a) 8.8 eV, (b) 10.3 eV, and (c) 10.0 eV, showing increasing disorder with formation energy. Large spheres denote atoms that are displaced by more than 10% of a bond length from their equilibrium positions.

There are two possible sources of entropic stabilization of the Humble/Arai configuration. The first is configurational; this structure possesses D_{2d} (four-fold rotational) symmetry, while the other configurations generally do not (see panels (b) and (c) for examples). However, this factor alone is insufficient to explain the peak in Figure 2 because the "excess" probability (i.e. the height of the peak relative to the overall trend) is almost two orders-of-magnitude. The only other possible factor is vibrational entropy. This was computed within the Quasi-Harmonic Approximation (8) for a large number of configurations of the 4-interstitial and 5-interstitial clusters, as shown in Figure 4. Results for two vacancy clusters (sizes 6 and 10) also are included for comparison.

The vibrational entropy of formation for all the clusters considered in this analysis generally is seen to increase approximately linearly with formation energy. This behavior can be attributed to the fact that higher energy configurations tend to be more spatially extended and involve more displaced atoms. As a result, more vibrational modes are likely to be introduced into the system, increasing the total vibrational entropy. While the vibrational entropies of vacancy and self-interstitial clusters both appear to exhibit this trend, vacancy clusters show a clearer linear trend, while the interstitial cluster data are substantially more scattered. One possible reason for this is that self-interstitial clusters tend to possess more internal structure, which can lead to larger fluctuations in the vibrational entropy of energetically neighboring configurations.

Most significantly, however, is that the anomalously large vibrational entropy of the 4-interstitial Humble/Arai configuration can be seen clearly in Figure 4(c) (denoted by the large circle symbol), which readily accounts for the excess probability associated with that particular structure relative to its energetically neighboring structures: ~5 k_B of additional vibrational entropy relative to energetically neighboring configurations increases the probability of observing this structure by exp(5)~O(10^2), fully accounting for the anomalous peak in the PDF of the 4-interstital cluster.

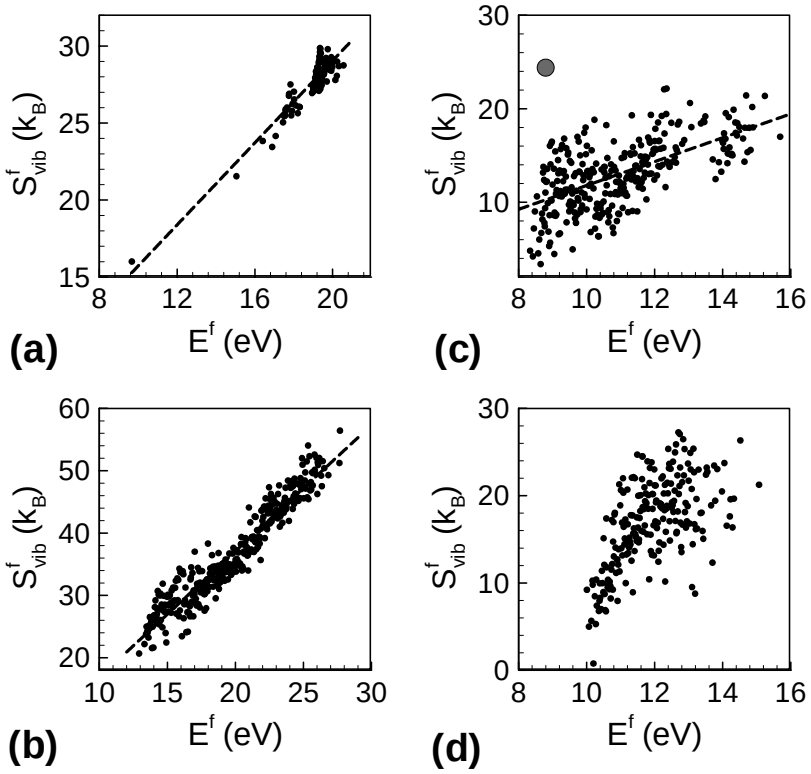


Figure 4. Vibrational entropy of formation as a function of formation energy for various vacancy and self-interstitial clusters. (a) 6-vacancy cluster, (b) 10-vacancy cluster (c) 4-interstitial cluster (Humble/Arai configuration, with its anomalously large vibrational entropy is highlighted with large circle), and (d) 5-interstitial cluster.

The very high vibrational entropy of the Humble/Arai structure has several important implications. First, it confers additional stability to the 4-interstitial cluster relative to neighboring sizes, which is likely to affect the dynamics of self-interstitial cluster growth and dissolution during annealing. This feature of the 4-interstitial cluster is consistent with previous literature studies in which inverse modeling of boron spike broadening predicted that the 4-interstitial and 8-interstitial clusters are particularly stable (17). Although not discussed here, a highly favorable configuration of the 8-interstitial cluster is comprised of two adjacent 4-interstitial Humble/Arai building blocks (18,19) which also possesses high vibrational entropy. Second, the entropic source of this stability implies that the effect persists at elevated temperature and thus is still relevant at typical processing temperatures. This is in stark contrast to the vacancy case where the particular stability of various sizes, which originates from bonding (enthalpic)

Ex. 1016, Page 993

considerations (see ref. (20) for details), is no longer relevant at high temperatures because it is effectively suppressed by entropic factors. One example of how the entropically stabilized 4-interstitial cluster can affect size distribution evolution during cluster growth or dissolution is presented in Figure 5, which shows the mean-square displacement of a 4-interstitial cluster as a function time at 1200K as obtained from direct MD simulation. The staircase-like behavior reflects configurational switching between the Humble/Arai configuration (immobile state) and all other disordered configurations (mobile states). As the temperature is lowered, the time spent in the immobile Humble/Arai configuration tends to increase, which renders the 4-interstitial cluster a potential bottleneck in cluster growth during wafer annealing or crystal growth under interstitial-rich conditions.

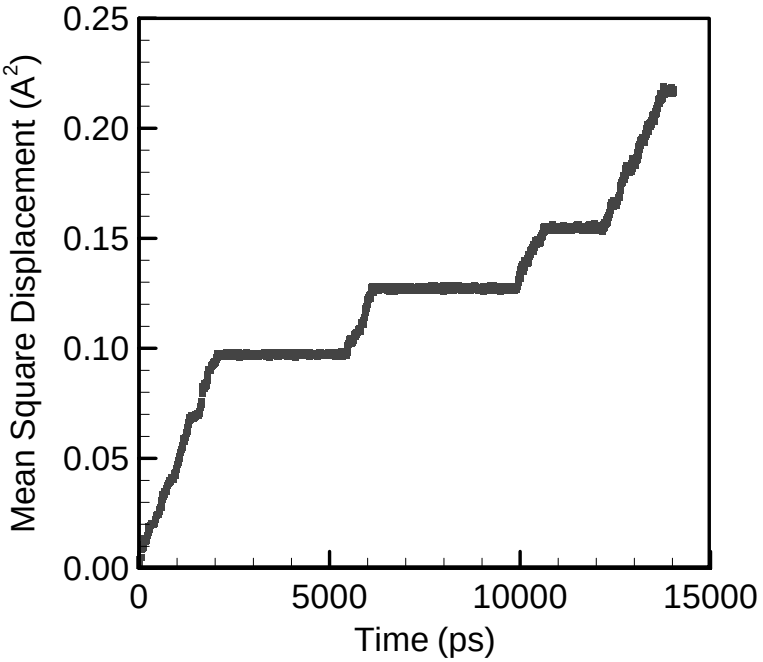


Figure 5. Mean-square displacement of 4-interstitial cluster as a function of time obtained by EDIP-MD simulation at 1200K. Flat sections correspond to the presence of the Humble/Arai configuration.

Conclusions

The thermodynamics of native point defect clusters in silicon are shown to be strongly influenced by both configurational and vibrational entropy. The existence of a large number of mechanically stable structures (i.e. a rough potential energy landscape) reduces the total formation free energy of defect clusters relative to the energetic ground state, particularly at the high temperatures encountered in crystal growth. As a result, the preferred small defect cluster configurations at elevated temperature tend to be extended, disordered structures that bear no resemblance to the minimum energy configuration. An exception to this is exhibited by some self-interstitial clusters, whose energetic ground-state configuration is additionally stabilized by high vibrational entropy. This entropic stabilization persists at high temperature and therefore is relevant to cluster size distribution evolution during processing.

The calculations presented here provide required inputs to continuum process models for crystal growth and wafer annealing. In the case of vacancy aggregation, inclusion of configurational entropy into vacancy cluster thermodynamic properties resolves an important discrepancy between previous cryogenic experimental measurements of the (111) surface energy (~1.25 J/m^2) and the much lower values required for continuum modeling of void formation during Czochralski crystal growth (~0.8-0.9 J/m^2 at 1100 °C). For self-interstitial clusters, the interplay between enthalpic and entropic stabilization of various configurations is possibly involved in the selection of the various defect structures observed experimentally, which range from small compact defects, to (113) planar defects, to (111) stacking faults.

Acknowledgments

Financial support of this work by the US National Science Foundation (CTS01-34418) and Siltronic AG is gratefully acknowledged.

References

1. V.V. Voronkov, *J. Crystal Growth*, **59**, 625 (1982).
2. D. J. Eaglesham, P. A. Stolk, H.-J. Gossmann, and J. M. Poate, Appl. Phys. Lett., 65, 2305 (1994).
3. T. Sinno, E. Dornberger, R. A. Brown, W. von Ammon, and F. Dupret, *Mater.Sci.& Eng.Rep*., **R28** 149 (2000).
4. M. S. Kulkarni, *J. Crystal Growth*, **303**, 438 (2007).
5. K. Sueoka, M. Akatsuka, M. Okui, and H. Katahama, *J. Electrochem. Soc*., **150**, 469 (2003).
6. S. Sastry, *Nature*, **409**, 164 (2001).
7. S. S. Kapur, M. Prasad, J. C. Crocker, T. Sinno, *Phys. Rev. B*, **72**, 014119 (2005).
8. W. G. Hoover, A. C. Hindmarsh, and B. L. Holian, *J. Chem. Phys*., **57**, 1980 (1972).
9. M. Z. Bazant, E. Kaxiras, and J. F. Justo, *Phys. Rev. B*, **56**, 8542 (1997).
10. F. H. Stillinger, Science, **267**, 1935 (1995).
11. T. A. Frewen, S. S. Kapur, W. Haeckl, W. von Ammon and T. Sinno, *J. Crystal Growth*, **279**, 258 (2005).
12. R. J. Jaccodine, *J. Electrochem. Soc*., **110**, 524 (1963).
13. J. J. Gilman, *J. Appl. Phys*., **31**, 3208 (1960).
14. F. Cristiano, N. Cherkashin, X. Hebras, P. Calvo, Y. Lamrani, E. Scheid, B. de Mauduit, B. Colombeau, W. Lerch, S. Paul, and A. Claverie, *Nuc. Instr. Meth. Phys. Res*., **216**, 46 (2004).
15. P. Humble, *Proc. Royal Soc. London A*, **381**, 65 (1982).
16. N. Arai, S. Takeda, and M. Kohyama, *Phys. Rev. Lett*., **78**, 4265 (1997).
17. N.E.B. Cowern, G. Mannino, P. A. Stolk, F. Roozeboom, H. G. A. Huizing, J. G. M. van Berkum, F Cristiano, A. Claverie, M. Jaraiz, *Phys. Rev. Lett.*, **82**, 4460 (1999).
18. S. S. Kapur and T. Sinno, *Appl. Phys. Lett*., **93**, 221911 (2008).
19. S. Lee and G. S. Hwang, *Phys. Rev. B*, **77**, 085210 (2008).
20. M. Prasad and T. Sinno, *Appl. Phys. Lett*., **80**, 1951 (2002).

ECS Transactions, 18 (1) 969-981 (2009)
10.1149/1.3096560 © The Electrochemical Society

Challenges in Wafering Processes for Semiconductor and Solar Wafers

Henry F. Erk[a]

[a] Corporate Technology, MEMC Electronic Materials, Inc., 501 Pearl Drive, St. Peters, MO 63376, USA

Abstract

The transition from 50 to 32nm silicon technologies will place strict requirements on semiconductor wafer flatness and nanotopography. Site flatness and shape capability at the edge must improve, as well as the edge characteristics (finish and shape) and quality. As diameter increases from 300 to 450mm, thickness does not scale with wafer diameter leading to a relatively less rigid wafer. This adds additional burden to the task of capability improvement.

To be competitive with alternative energy technologies solar wafer manufacture is focused on cost reduction and capability improvement to enable cell-processing optimization. The challenge is to cut thinner wafers with reduced depth of damage, lower TTV and with no micro-cracks. Slicing kerf reduction and abrasive / silicon recovery are options to further reduce cost.

Introduction

Silicon, second most abundant element of the planet, enabled the electronic revolution of the 20th century. Because of its unique physical and electrical properties and the ability to highly purify it, silicon was well suited to manufacture discrete transistors. As the electronics revolution progressed, discrete components reduced in size and became integrated onto a single chip. A wealth of knowledge was developed on how to refine, purify, crystallize, form, shape, etch and polish silicon. Because of this expertise, proposed alternative semiconductors with superior electrical properties (e.g. gallium arsenide and silicon carbide) have never replaced silicon. Economies of scale for device manufacturing pushed silicon wafers to larger and larger diameters. Round disks of silicon could be fabricated in large sizes with improved precision in flatness. It is highly likely in the near future, when circuits are formed with other more exotic materials, that the silicon wafer will remain as the handle, supporting the device layer; despite the fact that other materials with more suitable properties (e.g. thermal conductivity) exist.

Silicon wafer makers are constantly being challenged to improve. The challenges arise from three distinct motivations for change:

1. Reduce the cost.
2. Change wafer attribute (e.g. etched backside is polished)
3. Change specification. (e.g. site flatness, diameter).

Understanding the need for and benefit of cost reduction is straightforward. The other two motivations for change are more complex, often interrelated and ultimately can be

Ex. 1016, Page 997

merged into one. For example, a wafer attribute may need to change, to meet the requirement of a tightened specification.

The latter two challenges are driven by the industry strictly following Moore's Law, which states that the circuit density shall double every one to two years (1). The density increase is required for faster microprocessor speed, smaller memory size or greater image (pixel) resolution. To achieve this density increase, device Critical Dimension (CD) must shrink. To shrink CD, microcircuit fabrication unit operations, such as lithography, Chemical-Mechanical Planarization (CMP), plasma etch and Rapid Thermal Processing (RTP) are pushed to extremes of their windows of operation (2). Sometimes interactions between the process and the wafer occur. More often than not, the interactions cause yield loss and need to be eliminated. This is done by either tightening wafer specifications (e.g. reduce NanoTopology (NT)) or changing the wafer attribute (e.g. backside polishing). An example is the interaction of wafer topology with CMP film uniformity for Shallow Trench Isolation (STI) (3). Wafer front side topology has been found to influence CMP film removal uniformity. Wafer etching processes (acid etching) can influence backside topography (4). A final solution is to polish the backside (change the attribute from etched to polished) by Double-Sided Polishing (DSP) (5). This solution is effective and for all future wafer diameter increases, starting from 300mm, the wafer specification has been changed from "backside etched" to DSP.

Device manufacturers must continually reduce cost. Significant cost reduction can be realized if more useable devices are made per wafer. The useful area of a wafer can be enlarged by increasing the radius of the region that die are printed. This can be accomplished by reducing the Edge Exclusion (EE). Or more significantly, this can be accomplished by increasing the wafer size. Both approaches pose challenges to wafer manufacturing. The edge of the wafer is a discontinuity where effects such as polishing pad rebounding (6) or gas flow in epi (7) can cause flatness non-uniformity. Also, as the wafer diameter increases, the equipment necessary to manufacture the wafer becomes larger and more complex. Wafer handling becomes more difficult. To achieve good flatness / topology specifications as wafer diameter increases is a tough challenge as outlined in the most current technology roadmap (8).

Silicon has also found a role in the energy revolution. Again, because it is an easily machined semiconductor and can be fabricated at relatively low-cost. However, to reduce cost, more often than not the silicon is in the form of fragile Multi-crystalline (Mc) silicon, rather than single crystal. Also, to reduce cost, the thickness of the wafer is much less (c.a. 220μm for solar versus 650μm for semiconductor wafer of the same size). Slicing thin Mc silicon wafers poses a challenge. By far the greatest impact in cost reduction is the reduction of kerf loss in slicing. Since the wafers are so thin, a large percentage of the silicon is lost in the slicing step. Given the current economic situation, this has to be one area of focus.

This paper will briefly review these challenges with respect to silicon wafer manufacture and solar wafer manufacture.

Ex. 1016, Page 998

Semiconductor Wafer Challenges

Figure 1 shows a simplified, overall semiconductor wafer process flow. In this figure all ancillary operations; such as cleaning, laser marking, edge honing, thermal treatments, etc. are not shown. It should be noted that in this paper there is no discussion on wafer cleanliness (e.g. LLS and metals), epi and crystal properties. But there are certainly tough challenges in these areas as well; it is just beyond the scope of this paper to consider them here. The ground semiconductor-silicon ingot is cut into wafers in the slicing process. In the past ID blade slicing was used, but today a multi-wire saw is used to economically wafer the ingot. Kerf loss is significantly less and productivity significantly higher using the wire saw. The edges of the as-cut wafers are profiled to prevent wafer chipping. Sometimes there can be one or more additional edge honing operations that may follow somewhere downstream in the process flow. They are not included in this discussion, but can exist for some wafer manufacturers. Next, lapping or grinding mechanically flattens the wafers. This produces consistent thickness and flatness. Grinding is preferred, because it is more economical and cleaner. Then the wafers are chemically etched in either an acid bath or caustic bath to remove the mechanical damage. Acid etching produces a smooth surface, but with the potential for some flatness / topology features. Caustic etching produces a rough surface (which requires more polish stock removal), but preserves the flatness from lapping or grinding. Following etching the wafer front surface must be polished and usually the edge and backside. The order of edge and surface polishing can vary depending upon if the wafers are to be Single-Sided Polish (SSP) or DSP. For DSP, the wafers are held in carriers between the two polishing platens. The wafer edge rubs against carrier and can be damaged. This damage needs to be removed by Edge POLishing (EPOL); so EPOL follows rough polishing. On the other hand, since SSP wafers are polished without carriers, there is no edge damage. So these wafers can be edge polished before rough polishing; since it is easier to do so (no need to protect or clean the polished face from EPOL slurry, since it is not yet polished). This is why figure 1 shows an EPOL step preceding rough polishing and an EPOL step following rough polishing. Then, the wafers are Finish POLished (FPOL), generally using single-sided CMP tools. The wafering process stops here, but there will be cleaning inspection, perhaps epi deposition or other thermal treatments, such as annealing, and packaging before the wafer is shipped to the fab.

Cost Reduction. The challenge of cost reduction basically translates into systematically reducing kerf loss and process removals at each stage of the operation to the safe minimum level. By reducing kerf loss in slicing and reducing removals in lapping / grinding, etching, and the polishing steps; more wafers per kg. of silicon can be produced. Other means of achieving cost reduction is to increase throughput or simplify / reduce process steps. To effectively reduce removals, increase throughput, or reduce process steps; detailed process models are required. Models are used to simulate and subsequently improve processes leading to both reduced removals and improved capability. For example, wiresaw modeling has shown how warp evolves (9). Ingot orientation can also affect slicing warp and damage depth (10). For wafer grinding subsurface damage depth can be estimated (11) and an understanding of the effect of kinematics used to improve shape. This leads to a potential to reduce removal as well as a potential to improve capability (12). For acid etching the effect of mass transfer and bubble masking can influence wafer topography and lead to high stock removal

Ex. 1016, Page 999

requirements in polishing (4). Modeling of many wafer processing operations, as well as other operations, such as epi, crystal growth, directional solidification, SOI and new materials, and polysilicon production forms the basis for rapid advancement in the area of cost reduction as well as capability improvement (13).

Wafer Attribute Change. As the industry moves below the 65nm node immersion lithography was more frequently employed. With immersion lithography wafer edge quality becomes increasingly important. During immersion litho the lens moves rapidly over the wafer surface with a fluid filling the gap between the lens and the surface. The fluid can undergo turbulent flow with capillary forces on the trailing edge of the meniscus peeling off particles from the edge of the wafer (14) (15). Edge flakes (residues of surface films on the edge) may be transported by the turbulent flow of the immersion fluid from the edge inwards. This can lead to particle-associated defects on interior die. To prevent this from happening wafer edge quality must be very good with no residual etch pits, scratches and roughness. It should be mentioned that the level of surface finish that is required here goes beyond the original definition of "edge polish." Edge contour must be precisely defined so that edge film flaking is prevented, since the wafer edges may be cleaned / polished several times in the fab. Edge Bead Removal (EBR) processes in the fab require that the finished edge from the wafer supplier be precisely specified. Thus, the edge contour / shape must be more carefully controlled. Also, the wafer must be more carefully handled throughout the entire process to prevent chipping and scratching.

Incidentally, with immersion litho, there is also some evidence that certain non-uniform wafer topology can negatively impact the exposure quality (16). This amplifies the importance of NT specifications.

Another reason that edge quality is becoming critical is related to the more aggressive in-fab RTP. As CD decreases the allowable thermal budget also decreases. The time at temperature must be reduced to reduce dopant redistribution over smaller length scales. The temperature must be increased to affect the (annealing) kinetics. The ramp rates increase to reduce cycle time. RTP processing ramp rates are made steeper (17). In this situation the wafer can be subjected to high thermal stress. To prevent wafer slip or violent breakage with high thermal stresses the wafer edge must be completely damage free. The final wafer edge finishing process and subsequent handling must address this issue. So now the wafer edge (following the same evolution as the wafer backside) is approaching the high level of precision and perfection of the device surface.

Wafer Specification Change As mentioned earlier, specifications continue to tighten as a consequence of the implementation of Moore's Law. Moore's Law requires that the device density doubles on fixed-area die in x year(s). On this fixed area the number of devices in any year y (N_y) must increase from the number of devices in the previous year (N_{y-1}) by the following expression:,

$$N_y = 2^{\frac{1}{x}} N_{y-1} \qquad [1]$$

The area occupied by a device is given by the square of the critical dimension (CD). In year y it is CD_y^2 and in year y-1 CD_{y-1}^2. The total area remains constant, so

$$N_y CD_y^2 = N_{y-1} CD_{y-1}^2 \qquad [2]$$

Substituting equation 1 into 2 and simplifying, the ratio $\frac{CD_y}{CD_{y-1}}$ is given by

$$\frac{CD_y}{CD_{y-1}} = \sqrt{\frac{1}{2^{\frac{1}{x}}}} = 2^{-\frac{1}{2x}} \qquad [3]$$

If the specifications scale with the size of the CD and follow Moore's Law, the ratio of the specification for a year to that of the previous year should follow the form given by equation 3. Figure 2 shows a plot using data from the ITRS roadmaps (from 2000 to the current 2008 update) of the ratio of the target site flatness for a year to the target site flatness for the previous year. Also shown is the ratio of NT target for a year to the NT target for the previous year. The trend for both site flatness and NT trend closely around the x=3 line. It should be mentioned that beyond 2011, NT targets fall below ten, so a change of even one unit (the limiting resolution of the measurement) will move the ratio away from the trend. But in a following year the swing will eventually be in the opposite direction. Also the swing will increase in magnitude as the NT specification approaches unity.. Nonetheless, if these swings are averaged over several years the trend line is followed. When specifications for the process follow the trend shown in figure 2 the site flatness and NT must cumulatively decrease by a factor of $\frac{1}{\sqrt{2}}$ every 3 years.

Table 1A and 1B shows key wafer parameters as outlined by the update for 2008 ITRS roadmap (6). A change made from 2007 to 2008 was to back off of the EE specification from 1.5 to 2mm. Currently there is no manufacturable solution for 1.5mm EE. Also, sub-10nm NT and sub-20nm site flatness are problematical. Even remaining at the current diameter and moving to smaller CD poses challenges for site flatness and NT. The metrology originally developed to measure and map wafer flatness was based on capacitance measurements using an analysis spot size of about $20mm^2$. Today flatness and NT are measured using optical interferometery. Edge site flatness is no longer measured using the traditional rectangular device site map, but is now measured using a map of hypothetical sites in the form of small annular sectors at the edge. Free-mount CMP finish polishing using soft pads of very short planarization lengths (and non-optimized backing films and retaining rings) may be problematical as maximum deviations fall in the sub-20nm range (18). Alternative mounting / finishing process may have to be considered as edge site flatness specifications tighten.

One may recall the earlier discussions of

i. the increasing importance of edge quality and
ii. the fact that DSP can damage the edge of the wafer and edge polishing is required to remove this damage.

Edge polishing can cause edge shape degradation because the polishing extends inward as much as 1 to 2mm from the edge (19) (20). So as a consequence of improving edge quality (by EPOL after DSP) flatness degradation from 1 to over 2mm EE can occur.

Ex. 1016, Page 1001

This means that improvements in DSP (reducing or eliminating edge damage) and EPOL (further reducing edge shape degradation / roll-off) are required. It should be mentioned that the DSP carrier, made of (alloy or stainless) steel, usually has ring of a plastic material to cushion the wafer edge, minimizing this damage (21). The wear rate of the insert material is generally higher that that of the steel, so the thickness of the insert material, over many runs in DSP becomes much less than that of the steel carrier. The wear of the carrier insert is similar to the wear of the retaining ring used in CMP. However, the thickness of the carrier and insert can minimize pad de-compression / rebounding that can cause edge roll-off (22). So, if softer insert materials are used (to minimize edge damage) they will wear at a higher rate and be less effective at reducing pad rebounding / de-compression, so edge roll-off may increase.

To satisfy site flatness specifications for sub-20 nm range and NT specifications for sub-10 nm new wafer shaping / finishing technologies such as localized etching or localized deposition may have to be developed (8) (23). These techniques require that the wafer shape be precisely measured. Then regions that are thicker are selectively etched or regions that are thinner have additional silicon selectively deposited. The selective etching or deposition must produce a surface is identical to the non-processed area (or so close that only a small amount of FPOL is required). Also the process must not degrade the surface of the non-processed area. At the nanometer level, these can be particularly challenging tasks to perform at high throughput and low cost.

Meeting the ever-tightening flatness and NT specifications for even larger diameter wafers will be more challenging. On a relative basis, this wafer will have less flexural rigidity. The classical plate equation governs the small transverse displacement, *w*, of a thin plate under a distributed load, *p*:

$$\nabla^2 D \nabla^2 w = p \qquad [4]$$

The bending/flexural rigidity, *D*, of a plate is proportional to the cube of the thickness

$$D = \frac{Et^3}{12\left(1-\nu^2\right)} \qquad [5]$$

where *E* is Young's modulus and ν is Poisson's ratio (24). The maximum displacement, w_{max}, of a circular plate uniformly loaded and (simply) supported on the edge occurs at the center and is given by

$$w_{\max} = \frac{p r_0^4 \left(5+\nu\right)}{64 D \left(1+\nu\right)} \qquad [6]$$

where r_0 is the radius of the disk. As presented by Watanabe and Kramer (25) the scaling factor, based on the wafer sag model, can be defined as the ratio $\frac{r_0^4}{t^3}$. The ratio $\frac{d_0^4}{t^3}$ (differing by a factor of 1/16), where d_0 is the wafer diameter can also be used. Figure 3 shows a plot of the ratio $\frac{d_0^4}{t^3}$ versus the diameter, d_0. This figure illustrates that as the wafer diameter increases, the scaling factor is not constant, but actually increases

significantly. Since the ratio $\frac{d_0^4}{t^3}$ increases, this leads to an increase in w_{max}: i.e. a decrease in flexural rigidity. Incidentally, it is interesting to note that a parabolic curve with intercept through the origin can be used to fit the data.

Concern of wafer sagging in a horizontal orientation, such as in a furnace or in a cassette is relevant; however, as this ratio increases, the wafer can also exhibit a (relatively) larger deformation as it is being made. Portions of the wafer may flex more readily in a double-sided tool, such as a lapper or DSP polisher. The wafer may more easily flex under the (higher) loads in simultaneous double-sided grinding. This flexing can lead to shape degradation. Figure 4 illustrates the impact of wafer flexing in grinding, creating NT (26). Thus, it may be more difficult to meet flatness or NT specifications. It may be difficult to meet reduced EE edge site flatness specifications. This is because of the wafer edge flexing at lower length scales under loads: at constant pressure the load increases as the square of wafer radius, while the circumference only by the radius.

When the effective rigidity of the wafer decreases, pressure may need to be reduced to decrease the load. Reducing the load reduces deformation that results in improved process capability. For example, consider DSP. If a DSP is operated at the same pressure the load per carrier (assuming same number of wafers per carrier) will increase as the square of wafer diameter. As mentioned earlier, the use of DSP carriers to control pad rebounding / de-compression and improve edge flatness is well known (22). However, the carrier thickness is limited by the wafer thickness. So with a larger but relatively thinner wafer, the larger carriers are also relatively thinner. These carriers are subjected to higher load and may flex more. This holds true for either the planetary DSP (22) or the oscillating DSP (20). It may be necessary to reduce the pressure to reduce deformation. This argument applies even more for lapping where, to prevent formation of burs on the edges of the lapping plates, the wafers overhang the plates. Notwithstanding, the machining rate for mechanical and chemo-mechanical processes generally follows Preston's law (27) (28). As machining pressure is decreased, material removal rate decreases. Thus, throughput is reduced, which leads to a higher manufacturing cost because more capital equipment is required. Similar arguments apply at FPOL where the edge of the wafer can flex between the backing film and polishing pad.

Finally, as wafer diameter increases there is the need to scale-up the processing equipment. In going from 300 to 450mm, not much of the existing wafer processing equipment can be used. If older equipment is used, the throughput will be reduced, sometimes by as much as a factor of 3. New capital equipment must be purchased. In many cases the larger equipment must operate at somewhat reduced loads to minimize wafer or even tool deformation. Technologically, scale-up of the processing equipment can lead to issues of machine rigidity (larger, more massive bases) or temperature stability (more heat generation and more length to expand when temperature changes). It has been well established that the temperature rise in wire slicing, due to heat generation for the cutting process, will cause degradation in mechanical shape (warp) due to thermal expansion (29). For wire slicing, at the same plunge rate profile, cutting forces will increase by 50 %. This may lead to a degradation in shape). Or, the profiled plunge rate will have to be reduced. Thus, either longer spools of wire will be required, or the wire wear will increase because each segment of wire will have to do more cutting. Increased wire wear leads to a degradation in TTV and warp. As the wafer diameter increases the rate of heat generation will increase by a factor of 1.5. Thermal expansion with higher

heat loads in larger equipment will have to be managed. Possibly new approaches, such as using diamond abrasives, diamond-coated wire, or ingot rocking will be required.

Solar Wafer Challenges

Solar wafers pose unique challenges as well. The overall wafer process flow is relatively simple: rectangular bricks of silicon are sliced into thin squares of silicon, cleaned and sent to the cell manufacturer. The wafers produced in this process are thin: a 156mm square wafer is typically about 220 +/- 30μm thick.

Table 2 summarizes the solar wafer parametric roadmap proposed by NREL (30). In the long term, silicon material consumption per watt of solar energy produced needs to decrease. Also, just as in the semiconductor industry, the size of the wafer needs to increase to take advantage of economies of scale. But specifications tighten as wafer thickness is reduced.

Cost Reduction. Since solar wafers are so thin, it is easy to see that a considerable material yield loss occurs in slicing as kerf loss. With the thickness of the wire plus twice the abrasive mean particle size approaching the thickness of the wafer, yield in slicing as based on silicon input is not much greater than 50 % to 55%.

As in the case of semiconductor wafer slicing, a good physical understanding of the sawing process and good modeling are required to reduce wire thickness and abrasive size so that kerf loss can be reduced. And as this happens, TTV should be preserved. Defects such as deep wiremarks or micro-cracks need to be eliminated. Micro-cracks can lead to downstream wafer breakage.

Another approach to cost reduction is to reclaim the silicon kerf from the saw slurry exhaust. This is a particularly challenging concept, since the exhaust slurry not only contains silicon fines, but also metal fines (iron, copper and tin from the brass-coated wire), abrasive and organic compounds (either oil or propylene glycol). It is also desired to recover the abrasive, since it is costly (31). Abrasive recovery for semiconductor wafer slicing is routine today (32,33). Recovery of silicon kerf that is clean enough for solar wafer manufacture is more difficult, but progress is being made (34).

Wafer Specification Change. Table 2 shows a target of increasing wafer size. An increase in wafer size along with a reduction in wafer thickness for Mc silicon will pose challenges, since the material more susceptible to flexing and breakage. TTV needs to scale with thickness. This is because as the wafer becomes thinner, cell performance becomes a strong function of wafer thickness (35). So the TTV specification should reduce by about ½ the current specification (viz. 15μm from 30μm). In addition the Mc silicon needs to be crack free. Thinner wafers flex more and Mc silicon has a higher propensity to develop micro-cracks as it is flexed. There may be opportunities for utilizing new technologies, such as diamond abrasives, diamond-coated wire or EDM. However, those technologies need to be cost competitive with the conventional process.

The first step in solar cell fabrication is a stress-relief etch / texture etch. The etching process is hindered by the fact that saw damage of ~ 5μm must be removed. If the depth of damage is reduced to ~ 0.2μm the need for stress-relief etch would be eliminated (36). The texture etch (if necessary) of wafers with reduced damage depth would be far simpler

and has the potential to produce a surface of lower reflectance. The solar cell efficiency would be improved because more photons are captured.

Conclusions

Semiconductor wafer manufacturing always faces the constant challenges of cost reduction. Also, as the industry moves to more sophisticated processing many features of the wafer are no longer acceptable. Earliest wafers were bright etched with no edge rounding and not polished. Then, wafers were polished on one side, Soon afterwards, they were edge rounded. Next, wafers were polished on both sides and around the edge. Today, there is even more focus on edge quality. As the industry follows Moore's Law, there is continued emphasis on wafer flatness, especially at the edge. Future specifications for edge site flatness and NT will push wafer shape to a new dimension. This also becomes even more challenging as the wafer diameter increases and the wafer rigidity decreases. In the future wafer flatness and NT specifications will approach those of optical components which are much thicker than a wafer. High throughput processing while maintaining wafer rigidity is essential.

Solar wafer manufacturing is driven by cost. Achieve a doubling of the surface area with only ½ the thickness of today's wafers, ½ TTV and with reduced kerf loss. Improve quality by eliminating micro-cracks. Reduce slicing damage depth and eliminate the need for stress-relief etch and enable further optimization by cell manufactures of the wafer texture.

Acknowledgments

The critical inputs of Roland Vandamme, Sumeet Bhagavat, Timm Baars, Marco Maffe, Mike Seacrist, Bruce Kellerman, Graham Fisher, Bob Standley and Milind Kulkarni of MEMC are gratefully acknowledged.

References

1. G. E. Moore, *Electronics*, **38**, 8 (1965).
2. G. E. Moore, *Proc SPIE.*, **2437**, (1995).
3. A. Grenville, B. Coombs, J. Hutchinson, K. Kuhn, D. Miller, and P. M. Trocollo, Proc. *SPIE: Optical Microlithography XIII Proceedings*, pp.452-459 (2000).
4. M. S. Kulkarni and H. F. Erk, *J. Electrochem Soc.*, **147**, 1, pp. 176-188 (2000).
5. H. F. Erk, "*Wafer Nanotopography*," http://www.memc.com/t-wafer-nanotopography.asp (2005).
6. Yun-Biao Xin, *Int. J. Manuf. Sci. Tech.*, **1**, 2, (2000).
7. J. A. Pitney and S. Kommu, http://www.memc.com/PDF_Files/Key_Topics_PDF_Folder/2008-ECS-Oct-2008-abstract-Pitney.pdf. (2008).
8. *International Technology Roadmap for Semiconductors, 2008 Update Front End Processes*, http://www.itrs.net/Links/2008ITRS/Home2008.htm. (2008).
9. P. Gupta and M. S. Kulkarni, *ECS Trans.*, **2 (2)**, pp. 123-134 (2006).
10. S. Bhagavat and I. Kao, *International Journal of Machine Tool & Manufacture.* **46**, 5, pp. 531-541 (2005).

11. Z. J. Pei, S. Billingsley and S. Miura, *International Journal of Tools and Manufacture*, **39**, pp. 1103-1116 (1999).
12. G. J. Pietsch and M. Kerstan, *Precision Engineering*, **29**, pp. 189-196 (2005).
13. M. S. Kulkarni, The *5th International Symposium on Advanced Science and Technology of Silicon Materials (JSPS Si Symposium)*, http://fukuoka.riam.kyushu-u.ac.jp/hawaii2008/pdf/A-02.pdf (2008)
14. M. Kocsis, D. et al.. "Immersion Specific Defect Mechanisms: Findings and Recommendations for their control," *Proc. SPIE: Optical Microlithography XIX*,, **6154** (2006).
15. I. Pollentier, et al. "Innovative metrology for wafer edge defectivity in immersion lithography," *Proc. SPIE: Metrology, Inspection and Process Control for Microlithography XXI*, **6518** (2007).
16. A. Erdmann, P. Evanschitzky and P. De Bisschop, *Proc. SPIE: Optical Microlithography XVIII*, **5754**, (2005)
17. P. Ramanarayanan, K. Knutson, J. Hwang, J. Leanord, and S. Govindaraju, US Patent Application 20080242117 A1 (2008).
18. Z. Zhang, *Ph. D. Dissertation, Kansas State University*, pp. 41-56 (2007).
19. K. Mizushima, U.S Patent 6,962,521, B2 (2005)
20. T. Kato et al., U.S. Patent Application 2008/0096474 (2008).
21. http://www.prhoffman.com/literature/Insert_Carriers.pdf.
22. G. Wenski, T. Altmann, W. Winkler, G. Heier, G. Hölker, *Mat. Sci. Semicon. Proc.*, **5**, pp. 375-380 (2003).
23. M. Yanagisawa and T. Okaya, US Patent 6987701B2
24. http://www.efunda.com/formulae/solid_mechanics/plates/thexory.cfm.
25. M. Watanabe and S. Kramer, *Interface*, **15,** 4, pp. 28-31 (2006).
26. R. Vandamme and M. Bhagavat, *US Patent Application 2007/0179659 A1*, (2007).
27. J. W. Preston, *Trans. Opt. Soc.*, **23**, pp. 141-164 (1927).
28. N. J. Brown, P.C. Baker, R.T. Maney, M. Fouts, **Proc.** *SPIE: Contemporary Methods of Optical Fabrication* ,**306**, 16 (1981).
29. T. Yamada, M. Fukunaga, T. Ichikawa, K. Furuno, K. Makino, and A. Yokoyama., *Theoretical and Applied Mechanics*, **51**, pp. 251–258 (2002).
30. *National Solar Technology Roadmap: Wafer-Silicon PV*, http://www1.eere.energy.gov/solar/solar_america/pdfs/41733.pdf.
31. K. Kajimoto, *US Patent 6,929,537 B2*, (2005).
32. C. Zavattari and G. Fragiacomo, *US Patent 6,231,628 B2*, (2001).
33. C. Zavattari, G. Fragiacomo and E. Portaluppi, *US Patent 7,223,344 B2*, (2007).
34. T. Y. Wang, Y. C. Lin, C. Y. Tai, C.C. Fei, M. Y. Tseng and C. W. Lan, *Prog. Photovolt: Res. Apl.*, http://dx.doi.org/10.1002/pip.863 (2008).
35. W. P. Mulligan, M. A. Carandang, M. Dawson, D. M. De Ceuster. C. N. Stone and R. M. Swanson, "*Reducing Silicon Consumption By Leveraging Cell Efficiency,*" 21st European Photovoltaic Solar Energy Conference, Dreseden, September, 2006 http://www.sunpowercorp.com/Smarter-Solar/The-SunPower-Advantage/~/media/Downloads/smarter_solar/sunpower_dresden_paper.ashx. (2006).
36. J. Wohlgemuth and S.P. Shea, "*Large- Scale PV Module Manufacturing Using Ultra-thin Polycrystalline Silicon Solar Cells, Annual Subcontract Report 1 April 2002 – 30 September 2003,*" http://www.nrel.gov/docs/fy04osti/35887.pdf (2004).

Figures

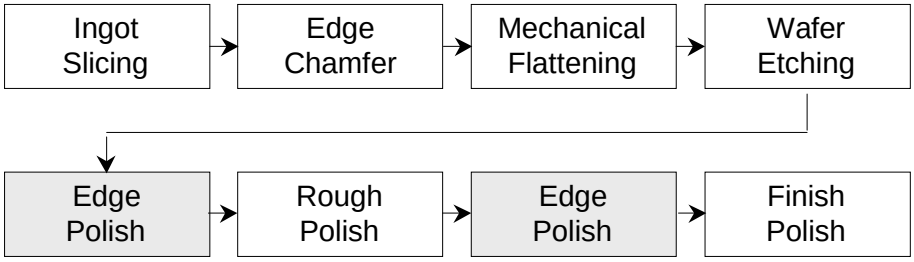


Figure 1. Semiconductor wafering process flow.

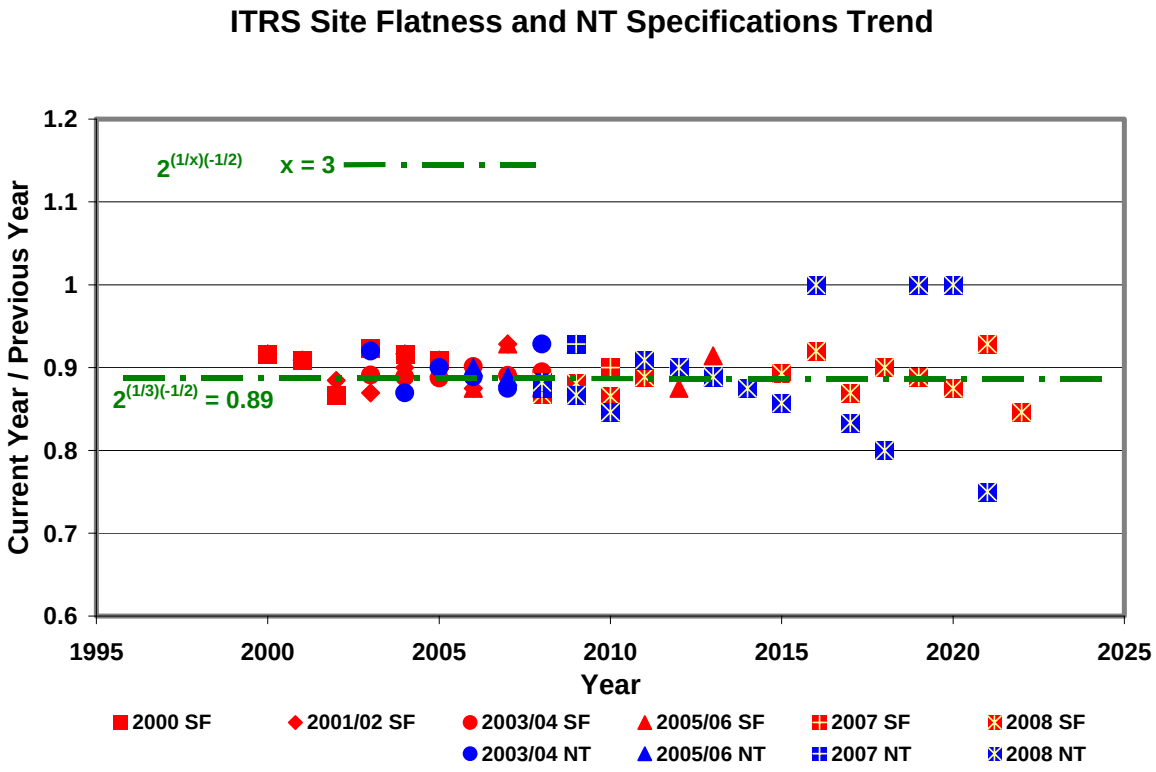


Figure 2. ITRS site flatness and NT specification trend. Year in the legend is the year of the ITRS roadmap or update. Abscissa is the ratio of the specification (Site Flatness = SF or NT) for the year on the ordinate divided by the previous year.

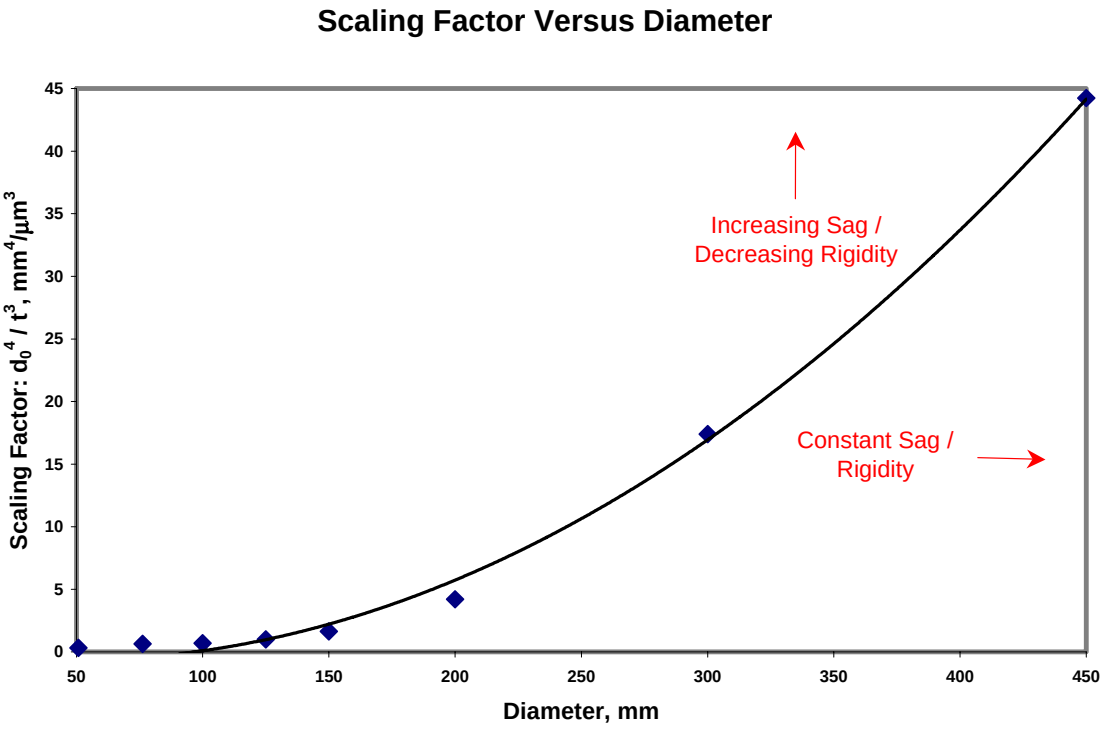
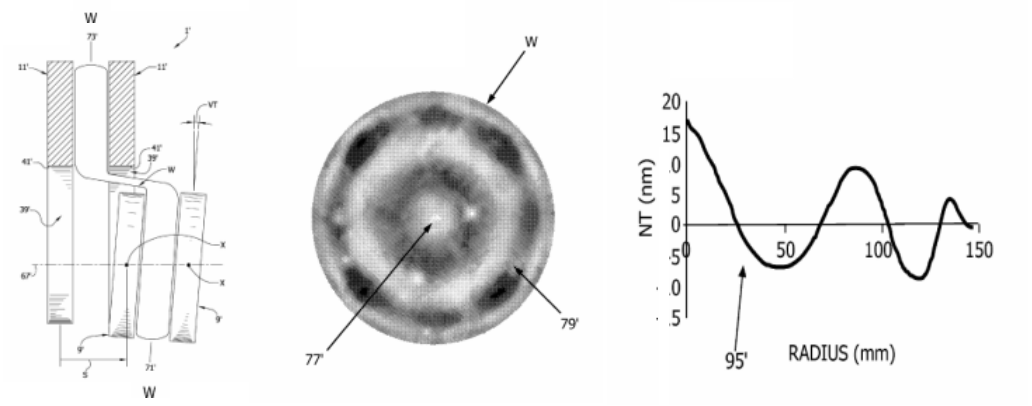


Figure 3. Wafer-sag model scaling factor versus wafer diameter.



3A 3B 3C

Figure 4. Wafer deformation in simultaneous double-sided grinding can effect wafer shape (26). Fig. 3A: exaggerated wafer deformation. Fig 3B: resulting NT map. Fig 3C: resulting PV radial-averaged NT scan.

Tables

Table 1A. 2008 Update ITRS roadmap for starting materials technology requirements (short term) (8)

Year	2007	2008	2009	2010	2011	2012	2013	2014
Wafer Diameter, mm	300	300	300	300	300	450	450	450
Edge exclusion, mm	2	2	2	2	2	2	2	2
Site Flatness: SFQR 26mm X 8mm	68	59	52	45	40	36	32	28
Nanotopography, P-V for 2mm spot size, nm	17	15	13	11	10	9	8	7

Table 1B. 2008 Update ITRS roadmap for starting materials technology requirements (long term) (8).

Year	2015	2016	2017	2018	2019	2020	2021	2022
Wafer Diameter, mm	450	450	450	450	450	450	450	450
Edge exclusion, mm	2	2	2	2	2	2	2	2
Site Flatness: SFQR 26mm X 8mm	25	23	20	18	16	14	13	11
Nanotopography, P-V for 2mm spot size, nm	6	6	5	4	4	4	3	3

Table 2. NREL national solar technology roadmap (30).

Year	2007	2015
Wire Sawing, $/W	25	15
Wafer Size, mm square	156	210
Wafer Thickness, mm	220	120

Ex. 1016, Page 1009

ECS Transactions, 18 (1) 983-988 (2009)
10.1149/1.3096561 © The Electrochemical Society

Prediction of Melt-Crystal Interface Shape and Melt Convection in a Large-scale CZ-Si Growth System Using RANS and LES Methods in Global Simulation.

Xin. Liu [a], Lijun. Liu [*,a], Zaoyang. Li [a], Yuan. Wang [a], and Koichi. Kakimoto [b]

[a] School of Energy and Power Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[b] Research Institute for Applied Mechanics, Kyushu University, Kasuga 816-8580, Japan

Abstract: Global simulations were performed for a large-scale CZ-Si growth system, in which three models were tested to predict the turbulent melt flow. The three turbulence models are the standard $k-\varepsilon$ turbulence model using wall functions at solid boundaries, a modified two-layer $k-\varepsilon$ turbulence model in RANS method and the standard Smagorinsky SGS model in LES method, respectively. The comparison of the simulation results obtained with these different models showed that each turbulence model has its own features in prediction of the melt convection and the melt-crystal interface shape in a large-scale CZ-Si growth system.

1. Introduction

The Czochralski (CZ) method has for several decades been the most widely used technique for growing high quality bulk crystals of silicon. The melt−crystal (m-c) interface shape and the melt flow motion have significant effects on the crystal quality in a CZ-Si growth process. As the grown crystal increases, the melt flow is mostly turbulent and turbulent fluctuation could be the dominant mechanism of momentum, energy and mass transport in the melt. Therefore, precisely modeling the turbulent melt flow and understanding the physical events in a large-scale CZ-Si growth system becomes essential [1, 2].

Over the last decade, there are many works on numerical analyses of large-scale CZ-Si growth system using different simulation techniques, which can be generally divided in two types: local models [3-5] and global models [6-8]. In the local models, attention is put on the melt and crystal domains only. To precisely model the turbulent melt flow, many turbulence models were applied and investigated in the frame of local simulation, such as the $k-\varepsilon$ family models in the RANS (Reynolds-Averaged-Navier-Stokes) method [3,4] and different SGS (sub-grid stress) models in LES (Large-Eddy Simulation) method [5]. However, since the solution of the melt flow highly depends on its boundary conditions and the CZ-Si growth furnace is a strongly conjugated thermal system, global simulation, in which the thermal field of the entire furnace is solved in a coupled way, is necessary to provide a 'real' boundary condition for the melt flow solution [6]. Some excellent works [7, 8] have been done on application of different turbulence models to model the turbulent melt flow in a CZ-Si growth system in the frame of global simulation. The purpose of this paper is to evaluate some representative turbulence models in RANS and LES methods for prediction of the melt convection and m-c interface shape for a large-scale CZ-Si growth system.

[*] Corresponding author: tel./ fax: 86-29-82663443, email: ljliu@mail.xjtu.edu.cn

2. Model Formulation

In the global simulation, some major assumptions are applied: (1) the geometry of the furnace is axisymmetric; (2) the growth system is quasi-steady; (3) the effect of gas flow in the chamber is neglected. The governing equations of the melt convection in a crucible can be thus written as

$$\nabla \cdot \vec{V} = 0 \quad [1]$$

$$\rho \vec{V} \cdot \nabla \vec{V} = -\nabla p + \nabla \cdot \left[\mu_{eff} \left(\nabla \vec{V} + \nabla \vec{V}^T \right) \right] - \rho g \beta_T \left(T - T_m \right) \quad [2]$$

$$\rho c_p \vec{V} \cdot \nabla \vec{T} = \nabla \cdot \left(\lambda_{eff} \nabla \vec{T} \right) \quad [3]$$

where $\mu_{eff} = \mu + \mu_t$ and $\lambda_{eff} = \lambda + \mu_t / \sigma_T$. The turbulent viscosity μ_t is calculated through applying turbulence models in the solution of melt flow.

The details of the global modeling algorithm were published in previous research papers [6, 9].

2.1 Turbulence models in RANS method

For the $k-\varepsilon$ family models in RANS method, the turbulent viscosity is calculated as $\mu_t = C_\mu \rho k^2 / \varepsilon$, where the governing equations for the turbulent kinetic energy k and its dissipation rate ε are:

$$\rho \vec{V} \cdot \nabla k = \nabla k \cdot \left(\left(\mu + \mu_t / \sigma_k \right) \nabla k \right) + + \left(P + G - \rho \varepsilon \right) \quad [4]$$

$$\rho \vec{V} \cdot \nabla \varepsilon = \nabla \varepsilon \cdot \left(\left(\mu + \mu_t / \sigma_\varepsilon \right) \nabla \varepsilon \right) + C_1 \left(P + G \right) \varepsilon / k - C_2 \rho \varepsilon^2 / k \quad [5]$$

Here C_μ, C_1 , C_2 σ_k , σ_ε and σ_T are constant model parameters.

Two $k-\varepsilon$ family models were applied as treatment of the near-wall turbulent flow is concerned. The first one is the well-known standard $k-\varepsilon$ model with the wall function, which can be found in many literatures [3].

The second one is the modified $k-\varepsilon$ model with the two-layer model for the near wall flows [10]. This model consists of the standard $k-\varepsilon$ model for the bulk flow and a one-equation model for the near-wall regions. In the near-wall region, Eq. [4] is used to compute k. Then ε and μ_t are calculated as $\varepsilon = k^{3/2} / l_\varepsilon$ and $\mu_t = \rho C_\mu l_\mu k^{1/2}$, respectively. The length scales l_ε and l_μ are estimated from

$$l_\varepsilon = 2.55 y \left(1 - \exp\left(-0.263 y^* \right) \right), \; l_\mu = 2.55 y \left(1 - \exp\left(-0.016 y^* \right) \right) \quad [6]$$

where y is the ordinary distance to the wall and $y^* = 2.55 y k^{1/2} / \nu$. At walls, $k = 0$ holds.

2.2 SGS model in LES method

In LES method, the vector formulations of the filtered governing equations have the same form as those in RANS method. The eddy viscosity μ_t is given by the Smagorinsky SGS model modified to account for the damping wall effect as follows [11]

$$\mu_t = D_w \rho \left(C_s \Delta\right)^2 \sqrt{2\overline{S}_{ij} \cdot \overline{S}_{ij} + \beta_T \overline{g} \cdot \nabla T / \sigma_T} \qquad [7]$$

where Δ is the grid-filter width. The empirical constant C_s is 0.3. The subscale turbulent Prandtl number σ_T is 0.4. The damping factor is calculated by the modified Van Driest damping function $D_w = \left[1 - \exp\left(-y^+ / 25\right)^3\right]^{1/2}$.

3. Results and discussions

In the series of computations, the input parameters are kept identical for all cases. The crystal diameter is 203 mm. The diameter of the crucible inside wall is 546 mm. The melting point of silicon is 1685K. The crystal pulling rate is 0.88 mm/min. The rotation rates of crystal and crucible are 8 rpm and -9 rpm, respectively.

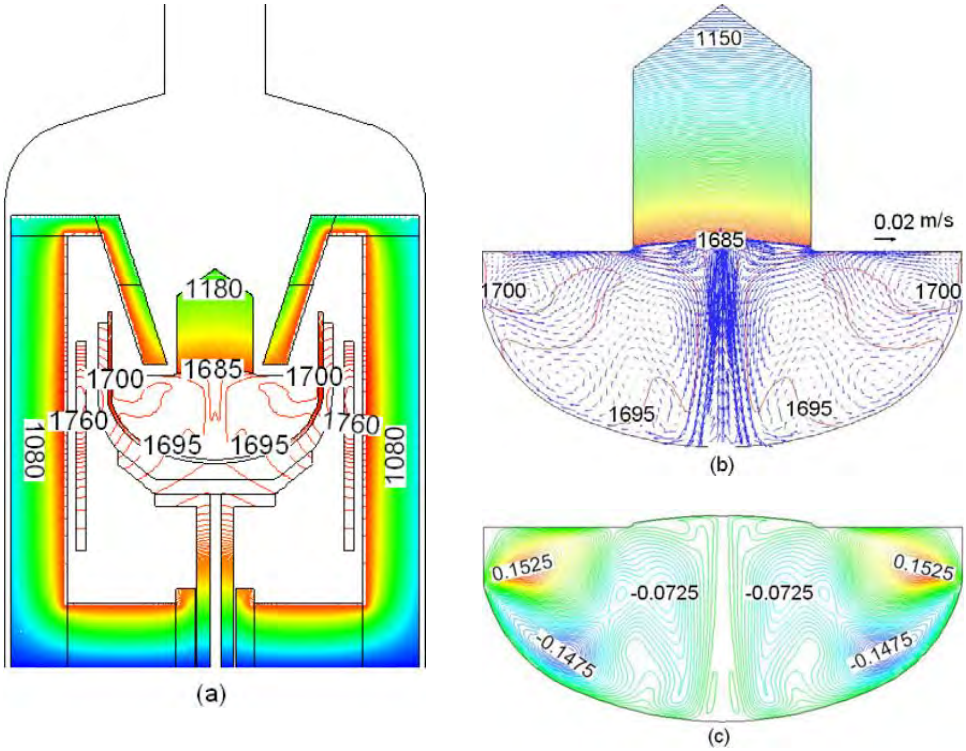


Figure 1. Thermal field and melt convection predicted with the laminar melt flow model. Isotherms of temperature are plotted every 5 K, interval of streamline function is 5×10^{-3} m^3/s: (a) global thermal field, (b) velocity field and temperature distribution in the m-c region, (c) contours of stream function in the melt.

First, we show the thermal field and melt convection predicted with the laminar melt flow model in Fig. 1. The temperature distribution in the entire furnace is presented in Fig. 1 (a). The insulating layer in the heat shields and gas-guide tube guarantees a reasonable hot zone in the m-c domain. Fig. 1 (b) shows the melt convection and thermal field in the m–c domain. The m-c interface corresponds to the isotherm of 1685 K, which is concave with respect to the crystal. The radial difference of temperature at the melt free surface is 15 K and the axial difference along the centerline is 10 K. The streamlines of the melt flow are plotted in Fig. 1 (c). We can notice that the flow pattern is chaotic and several vortexes were predicted. In order to precisely predict the heat transfer and m-c interface shape, turbulence transport must be considered.

Fig. 2 shows the results predicted with the standard $k-\varepsilon$ model with wall function applied at solid boundaries. Fig. 2(a) gives the melt velocity and thermal field. Fig. 2 (b) presents the stream function distribution of the melt flow. Here a cellular motion with the stream function $\psi > 0$ corresponds to a clockwise motion and that with $\psi < 0$ corresponds to a counter clock-wise motion. The balance between the centrifugal force and the pressure gradient leads to a radial motion towards the centerline under the crystal and the formation of a counter clockwise rotating vortex in the melt. The maximal radial difference of temperature is 30 K at the melt surface and the maximal axial difference at the centerline is about 15 K. The thermal field in the melt is more uniform than that predicted with the laminar flow model due to the effect of extra turbulent diffusion.

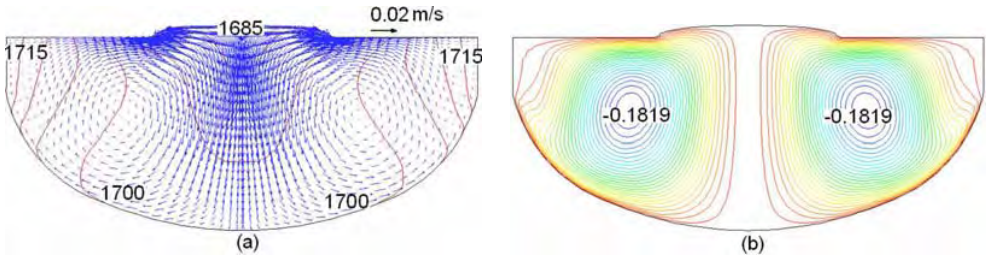


Figure 2. Melt convection and m-c interface shape predicted with the standard $k-\varepsilon$ turbulence model. Isotherms of temperature are plotted every 5 K and the interval of streamline function is 5×10^{-3} m^3/s: (a) velocity field and temperature distribution, (b) contours of stream function in the melt.

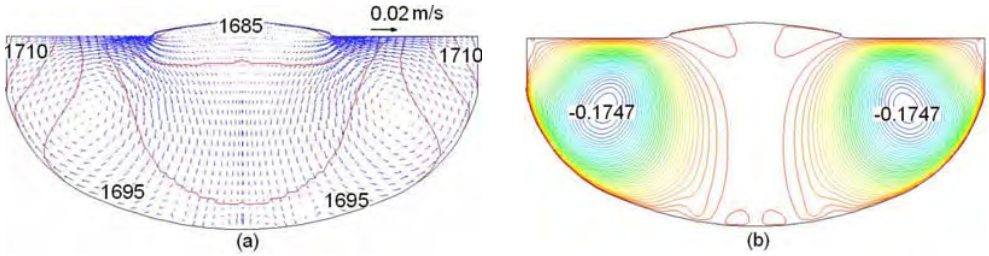


Figure 3. Melt convection and m-c interface shape predicted with the two-layer modified $k-\varepsilon$ turbulence model. Isotherms of Temperature are plotted every 5 K and the interval of streamline function is 5×10^{-3} m^3/s: (a) velocity field and temperature distribution of the melt, (b) contours of stream function in the melt.

Global simulation was also conducted by applying the two-layer modified $k-\varepsilon$

model in RANS method. The results are presented in Fig. 3. Because the modified $k-\varepsilon$ model differs only in the near-wall flow treatment comparing to the standard model, similar flow patterns distributions were noticed as comparing the results in Fig. 3 with those in Fig. 2.

Over the past decade, LES has been developed as an important tool to model turbulent flows. In our series of computation, we also used LES method to model the melt turbulent flow in the global simulations and the standard Smagorinsky model with wall damping function was applied in the LES method. The predicted results are presented in Fig 4. We can notice quite some small vortexes in the melt. The flow structure shows remarkable difference comparing to those predicted with the $k-\varepsilon$ family turbulence models in RANS method. However, the maximal radial and axial temperature differences in the melt are approximately equal with those predicted with RANS method.

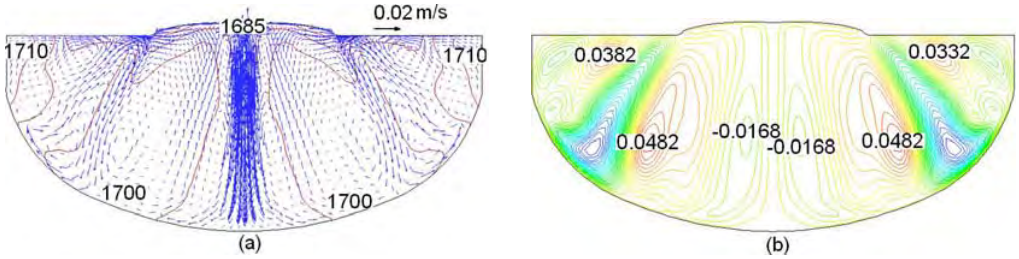


Figure 4. Melt convection and melt-crystal interface shape predicted with the standard Smagorinsky SGS model in LES method. Isotherms of temperature are plotted every 5 K and the interval of streamline function is 5×10^{-3} m^3/s: (a) velocity field and temperature distribution, (b) contours of stream function in the melt.

We compared the m-c interface shapes respectively predicted with different models in Fig. 5. The predicted interface shapes are concave to the crystal as observed in practice. The comparison shows little difference between those predicted with different models.

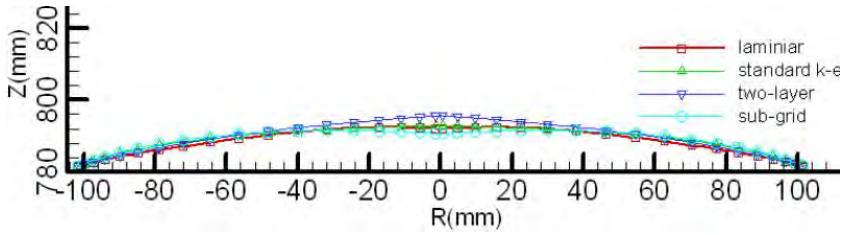


Figure 5. m-c interface shape predicted with different turbulence models

4. Conclusions

Two $k-\varepsilon$ family models of turbulence in RANS method and the standard Smagorinsky SGS model in LES method for the melt turbulent flow were applied and evaluated to predict the melt convection and the m-c interface shape in the frame of global simulation for a large-scale CZ-Si growth system. Some conclusions could be derived from the comparison of these simulation results. First, turbulence transport has significant influence on the melt convection and heat transfer. It must be taken into account in the global simulation. Second, the three turbulence models present tiny

distinction in predicting the m-c interface shape in a 2D global simulation. Third, the $k-\varepsilon$ family models in RANS method can qualitatively predict the melt convection with lower mesh requirement. But flow structure in small scales can not be resolved. Last, the LES method exhibits higher resolution of flow structure than RANS method.

Acknowledgments

This work was supported by Project 50876084 supported by NSFC and the project supported by Natural Science Basic Research Plan in Shaanxi Province of China (Program No. SJ08E$_2$01).

References

1. T. Sinno, E. Dornberger, W. von Ammon, R.A. Brown and F. Dupret, *Materials Science and Engineering: R: Reports*, 2000. 28(5-6): p. 149-198.
2. E. Tomzig, W. von Ammon, E. Dornberger, U. Lambert and W. Zulehner, *Microelectronic Engineering*, 1999. 45: p. 113-125.
3. B. Yu and H. Ozoe, *Journal of Materials Processing & Manufacturing Science*, 1999. 8(2): p. 162-185.
4. A. Lipchin and R.A. Brown, *Journal of Crystal Growth*, 1999. 205(1-2): p. 71-91.
5. I.Y. Evstratov, V.V. Kalaev, A.I. Zhmakin, Y.N. Makarov, A.G. Abramov, N.G. Ivanov, E.M. Smirnov, E. Dornberger, J. Virbulis, E. Tomzig, and W. von Ammon,*Journal of Crystal Growth*, 2001. 230(1-2): p. 22-29.
6. Lijun. Liu and K. Kakimoto, *International Journal of Heat and Mass Transfer*, 2005. 48(21-22): p. 4481-4491.
7. T.A. Kinney and R.A. Brown, *Journal of Crystal Growth*, 1993. 132(3-4): p. 551-574.
8. R. Assaker, N. van den Bogaert, and F. Dupret, *Journal of Crystal Growth*, 1997. 180(3-4): p. 450-460.
9. Lijun. Liu, T. Kitashima, and K. Kakimoto, *Journal of Crystal Growth*, 2005. 275(1-2): p. e2135-e2139.
10. H. Iacovides, B.E. Launder, and H.Y. Li, *International Journal of Heat and Fluid Flow*, 1996. 17(1): p. 22-33.
11. I.Y. Evstratov, V.V. Kalaev, A.I. Zhmakin, Y.N. Makarov, A.G. Abramov, N.G. Ivanov, E.M. Smirnov, E. Dornberger, J. Virbulis, E. Tomzig, and W. von Ammon, *Journal of Crystal Growth*, 2001. 230(1-2): p. 22-29.

ECS Transactions, 18 (1) 989-994 (2009)
10.1149/1.3096562 © The Electrochemical Society

Calibration of the Enthalpy-Porosity Based Method for CZ Silicon Growth

Bing Dai[a], Balaji Devulapalli[a], Prashant Gunjal[b], and Milind Kulkarni[a]

[a] MEMC Electronic Materials Inc., St. Peters, Missouri 63376, USA
[b] Tridiagonal Solutions, Pune, India

A global transport model for industrial Czochralski (CZ) silicon growth is reported in the paper. The model quantifies the temperature fields in all components of the growth system. The heat transport by the conduction, convection (melt and argon flows) and radiation are treated along with the momentum balance in the presence of the static magnetic fields. The crystallization process is modeled by an enthalpy-porosity method in which, the phase change occurs across a finite region defined by volumetric sink/source of heat (enthalpy) and momentum. The momentum sink/source is formulated based on the Darcy's law for porous medium, which is simply a numerical treatment for the isothermal solidification. The flow field in the Czochralski melt predicted by the global enthalpy-porosity based model is compared with that predicted by a more accurate simplified local model. Our studies indicate that the intensity of the flow induced by the presence of the crystal/melt interface is underestimated by the enthalpy-porosity based formulation.

Introduction

Early modeling studies of Czochralski (CZ) silicon growth date back to the 1980s (1, 2). An overwhelming majority of these studies treat the isothermal CZ silicon crystallization by the Stefan condition at the crystal/melt interface. The interface shape is determined by satisfying the Stefan condition along the isotherm at the melting temperature. A large number of numerical techniques were developed for the solution of this problem. The front-tracking techniques provide an accurate physical description of the isothermal phase change for the CZ silicon growth. The application of these techniques requires a careful development of tailor-made computational codes and computational effort. The front-tracking techniques involve direct treatment of two phases marked by a discontinuity at the interface. Numerical and physical approaches that quantify the solidification in one continuous domain without discontinuities have their own advantages over the formulations based on the Stefan condition.

The various daily engineering problems as well as the demand for rapid delivery of results in the manufacturing sector render the use of commercial software very attractive, especially to solve the flow dynamics in physical systems. These software are developed to solve general flow problems. We have carried out global simulations for the CZ silicon growth system at MEMC Electronic Materials (MEMC) in the framework of the popular flow simulator, Fluent. The CZ silicon crystallization is modeled by an enthalpy-porosity based method using a fixed mesh (3, 4). According to the enthalpy-porosity based methods, sink/source terms are introduced in the momentum and energy balance to account for the phase change. The enthalpy-porosity based methods accurately quantify

Ex. 1016, Page 1017

the solidification of mixtures over a finite temperature range; but they can only approximately quantify the isothermal phase change of monocomponent systems. Careful calibration of the enthalpy-porosity based methods to other physically sound formulations is thus valuable.

In this study, we compare various CZ melt flow fields predicted by the enthalpy-porosity based method with those predicted by a simple local model that describes the CZ melt confined within solid surfaces. The local model isolates the crystal and melt from the rest of the CZ silicon growth system. In the local model, the crystal/melt interface is treated as a rigid wall. The interfacial shape and the conditions at the boundaries of the local model are given by the global simulation. All other settings of the local model remain the same as the global simulation, which involves enthalpy-porosity based quantification of the phase change. The melt momentum balance in the enthalpy-porosity based global model and that in the local model are treated by the same turbulence approximations. Thus, the local model can be used to calibrate the global model, in a semi-quantitative sense. The comparison indicates that the enthalpy-porosity based global model is incapable of accurately capturing the momentum exchange between the crystal and melt. Moreover, the flow strength under the crystal is underestimated by the global model. The results of our study are described in detail in this paper.

Model Formulation

We have set up a global transport model for the industrial CZ silicon growth system. In the global formulation, the crystallization process is modeled by an enthalpy porosity method. Upon the convergence of the global simulation, the crystal and melt are separated from the rest system for the localized modeling. The global simulation provides the boundary conditions for the local model, including the melt/crystal interface shape.

In the global simulation, a quasi-steady-state is assumed because the timescale of momentum and heat transport are much shorter than the timescale of CZ growth. The CZ system is assumed to be symmetric around the growth axis of the crystal. The temperature field is computed in all components or bodies of the system. The components exchange energy by radiation and convection; the surfaces of all components are assumed to be diffuse and gray. The transparent or semi-transparent components are described by the discrete ordinates radiation model. The convective heat transfer is dominant in the fluid domains, i.e., silicon melt and argon gas. A two-equation k-ε turbulence model with the low-Reynolds number features is applied to study the melt flow dynamics. The silicon melt flow is influenced by the buoyancy driven by the temperature variations in the melt; the effect of buoyancy is modeled by the Boussinesq approximation. The melt flow is also influenced by the crucible and crystal rotations and the Marangoni effects, that arise because of the temperature dependence of the surface tension of the silicon melt. For simplicity, the thermal capillarity of the silicon melt/argon interface is neglected.

As discussed before, for the simulation of phase change in the global CZ system, we use the enthalpy-porosity based method instead of the front-tracking technique. The primary reason for this choice is its availability in the commercial software (Fluent) and the ease of its application using a fixed mesh. The enthalpy-porosity based approaches do not accurately capture the flow field induced by the interaction between the crystal and the melt, in monocomponent systems, which are better represented by the Stefan condition. It

Ex. 1016, Page 1018

is important to understand the key differences between the enthalpy-porosity based methods and the front-tracking methods based on the Stefan condition. The front-tracking technique treats the crystal and melt as separate phases or domains. The exchange of energy between these domains is described by the Stefan condition. For momentum balance no-slip condition is applied. Stefan condition results from the energy balance across the crystal/melt interface:

$$-K_c(\nabla T)_c \cdot \mathbf{n} = -K_m(\nabla T)_m \cdot \mathbf{n} + \rho_c L \mathbf{V} \cdot \mathbf{n} \quad [1]$$

Where K_c and K_m are the thermal conductivities of the crystal and the melt, respectively, L is the latent heat of silicon per unit mass, **V** is the crystal pull-rate and **n** is the unit vector normal to the interface. The interface shape is determined by satisfying the Stefan condition at the position of the isotherm at the melting temperature ($T_{int}=T_M$ where T_{int} is the interfacial temperature and T_M is the melting temperature) are satisfied.

In the framework of the enthalpy-porosity technique, a new quantity, the liquid fraction β is introduced, and for the isothermal solidification, β is defined as follows:

$$\beta = \begin{cases} 1 & T > T_M \\ 0 & T < T_M \end{cases} \quad [2]$$

where T_M is the melting temperature. The energy balance is written for the total enthalpy H, which is the sum of the sensible heat h and the latent heat, i.e, $H=h+\beta L$. This treatment introduces an additional volumetric sink/source term in the standard formulation for the sensible heat (h) balance to account for the phase change. For steady-state problems, the sink/source term is given by,

$$S_h = \nabla \cdot [\rho_c \mathbf{u} \beta L] \quad [3]$$

where **u** is the ensemble averaged velocity which is continuous throughout the crystal and melt. The liquid fraction β is updated by the following numerical scheme:

$$\beta_{i+1} = \beta_i + \omega c[T_i - T_M]/L \quad [4]$$

where c is the heat capacity, ω includes the relaxation factor and other numerical terms and the subscript stands for the iteration sequence, i.e., i+1 means the (i+1)th iteration. Equation (4) closely mimics Equation (2). It is noted that, β is continuous in this formulation but changes sharply across the narrow region of solidification.

Similar to the energy balance, additional sink/source term is introduced into the momentum balance,

$$S_M = A\frac{(1-\beta)^2}{(\beta^3+\varepsilon)}(\mathbf{u}-\mathbf{u}_c) \quad [5]$$

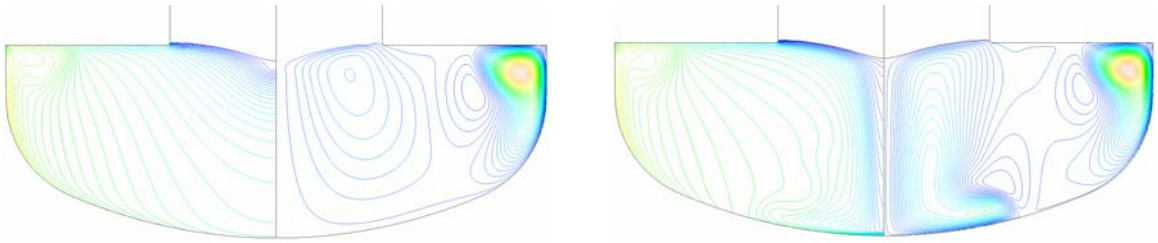
Where A is a constant (we set 10^{10} in the calculations), ε is a small number to prevent division by zero, and $\mathbf{u}_c$ is the bulk crystal velocity. With extremely large A, the sink/source term is dominant in the momentum balance and the fluid velocity is forced to be the same as the bulk crystal velocity in the crystal and in the crystal/melt transition region where the liquid fraction is close to zero. The momentum sink/source in Equation (5) is based on the Darcy's law for porous media. For the isothermal solidification in CZ silicon growth, this sink/source term is a numerical fix.

The CUSP magnetic field is applied to the crystal and melt that have finite electrical conductivity and the rest of the CZ system is treated as a perfect insulator. The electrical potential is solved in the crystal and melt. The electrical conductivity is defined as a linear function of the liquid fraction β, i.e., $\sigma = \beta\sigma_m + (1-\beta)\sigma_c$ where σ_m and σ_c are the electrical conductivities of pure silicon melt and crystal, respectively. The induced Lorentz force and Joule heat by the magnetic effects are introduced as additional sink/source terms into the momentum and energy balances, respectively (5).

For the local simulation, the crystal and melt the are separated from the global system. In the local model, the temperature profiles at outer walls of the crystal and melt are given by the global simulation based on the enthalpy-porosity approximation. The global simulation also provides the crystal/melt interface shape for the local model, in which the interface is treated as a rigid wall at the melting point. All other settings of the local model, i.e., the turbulence formulations and the magnetic fields, remain exactly the same as the global model.

Results and Discussion

In this section, we compare the results obtained by the global simulation based on the enthalpy-porosity methods to the results obtained by solving the local model. In all calculations, the crystal and crucible are rotate in the opposite directions. Different standard CUSP magnetic fields are applied and the CUSP location, or the location of zero magnetic field strength, can be varied by adjusting the current density in the coils. We set it either in the crystal (C1M0) or at the crystal/melt interface (C0M0).



a b

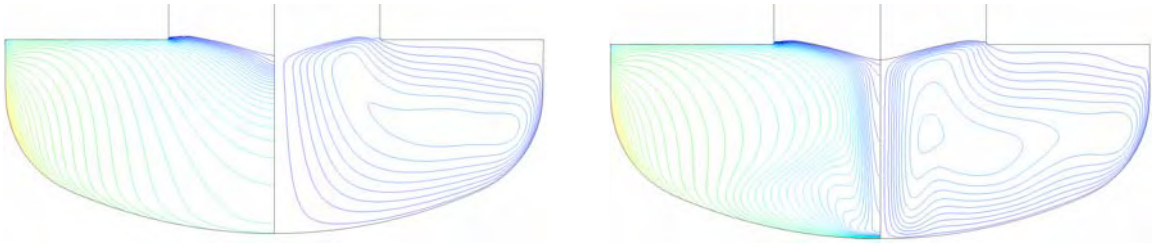
Figure 1. Temperature field (left) and stream lines (right) in the CZ melt without magnetic effects. a) global model, b) local model.

The results without magnetic effects are shown in Figure 1. A big counter-clock-wise flow cell forms near the crucible wall in both the global and local simulations driven by the buoyancy effect. The flow vortex generated by the buoyancy effect does not penetrate deeply into the melt because of the pressure gradient established to maintain the rigid-

body motion in the melt. As the melt flows downward in the buoyancy vortex, closer to the axis of symmetry, secondary or intermediate cells are formed. Under the crystal the melt flows toward the rotating axis because of the imbalance between the radial pressure gradient and the centrifugal force, induced by the rotation of the crystal/melt interface, which disturbs the rigid-body rotation of the melt. This is known as the Taylor-Proudman effect (6).

It is clear that the enthalpy-porosity based global model predicts a weaker flow below the crystal than the local simulation that accounts for the presence of a solid interface in contact with the melt. The flow under the crystal is much stronger in the local simulation. The maximum value of the stream function in the Taylor-Proudman vortex in the local simulation is twice of that in the global simulation (The isotherms and streamlines in all figures are plotted against the same scales). This affects the temperature field below the crystal.

In the local simulation, the crystal/melt interface is treated as a rigid wall with no-slip condition. In the global model, the bulk melt is not in the direct contact with the rigid interface. The bulk melt feels the presence of the crystal only through a narrow mushy zone, across which solidification occurs. The step change between the crystal and melt is approximated as a differentiable sharp change in the global formulation. Fine mesh in the vicinity of the crystal/melt interface improves the prediction of the global model. Modifications to the global model that mimic the boundary layer dynamics under the rotating crystal can also improve the predictions of the model.



a b

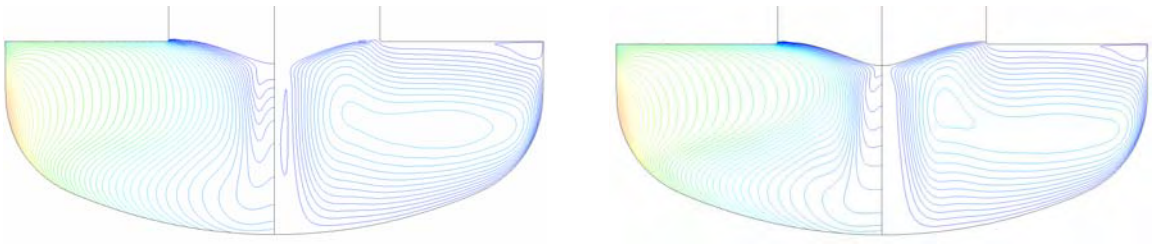
Figure 2. Temperature field (left) and stream lines (right) of the growth system in the CUSP magnetic field (C0M0); a) global model, b) local model.

The overall intensity of the flow is reduced in the presence of the CUSP magnetic field. The maximum value of the stream function in the presence of the magnetic fields is significantly lower (Figures 2 and 3) than that in the absence of the magnetic fields (Figure 1). The secondary cells formed in the absence of magnetic field disappear in the presence of the magnetic field, as shown in Figure 2 and 3. Only one counter-clock-wise flow vortex remains in the outer melt. The reduction in the intensity of the flow field reduces the heat transfer efficiency from the crucible wall to the crystal. To maintain the melting temperature at the junction of melt, crystal and argon, or the tri-junction point that defines the crystal radius, the heater power increases along with the maximum melt temperature, in the presence of the magnetic field.

In the enthalpy porosity based global simulation, the electrical potential is solved in the domains of crystal and melt. For the sake of simplicity, the local model treats crystal/melt interface as an insulating wall. Because of this approximation, a comparison of the global

Ex. 1016, Page 1021

model and the local model in the presence of the magnetic fields is only semi-quantitative. This approximation, however, is partly justified because of the low electrical conductivity of silicon crystal, which is around 50 times lower than that of silicon melt. The main difference in the predictions of the global model and the local model remains the flow field under the crystal as shown in Figures 2 and 3. The Taylor-Proudman effect is underestimated in the global simulation.



a b

Figure 3. Temperature field (left) and stream lines (right) of the growth system in the CUSP magnetic field (C1M0); a) global model, b) local model.

Concluding Remarks

We carry out a comparative study of an enthalpy-porosity based model and a model based on discontinuous solidification, for the industrial Czochralski (CZ) silicon growth. The entire global CZ growth system is simulated using the enthalpy-porosity based model but only the local melt and crystal domains are simulated using the discontinuous solidification model. Moreover, in the local model, the crystal/melt interface is treated as a rigid wall at the melting temperature. The comparisons indicate that the Taylor-Proudman effect is underestimated in the global simulation, both in presence and absence of the steady magnetic fields. The momentum exchange between the crystal and the melt is not accurately quantified by the enthalpy-porosity based model. Modifications to account for this approximation are required to accurately predict the effects of the thin boundary layer below the crystal.

References

1. P. A. Sackinger and R. A. Brown, *Int. J. Numer. Meth. Fluid*, **9**, 453 (1989).
2. J. J. Derby and R. A. Brown, *J. Cryst. Growth*, **87**, 251 (1988).
3. V. R. Voller, M. Cross and N. C. Markatos, *Int. J. Numer. Meth. Eng.*, **24**, 271 (1987).
4. V. R. Voller and C. PRAKASH, *Int. J. Heat Mass Transfer*, **30**, 1709 (1987).
5. P. A. Davidson, *An Introduction to Magnetohydrodynamics*, Cambridge University Press, Cambridge (2001).
6. H. P. Greenspan, *The Theory of Rotating Fluids*, Cambridge University Press, Cambridge (1968).

ECS Transactions, 18 (1) 995-1000 (2009)
10.1149/1.3096563 © The Electrochemical Society

Advances in the Understanding of Oxide Precipitate Nucleation in Silicon

G. Kissinger[1], D. Kot[1], V. Akhmetov[2], A. Sattler[3], T. Müller[3], W. von Ammon[3]

[1] IHP, Im Technologiepark 25, 15236 Frankfurt (Oder), Germany
[2] IHP/BTU Joint Lab, Universitätsplatz 3-4, 03044 Cottbus, Germany
[3] Siltronic AG, Hanns-Seidel-Platz 4, 81737 München, Germany

The influence of vacancy supersaturation installed by RTA pre-treatments in CZ silicon wafers on oxide precipitate nucleation was investigated in the temperature range 700-1000 °C. Precipitation is enhanced at 800 °C and increases with increasing vacancy concentration. Getter efficiency tests for Cu and Ni have shown that the threshold value of the normalized inner surface is shifting to higher values for increasing RTA temperature. This can be explained by a morphological change of the oxide precipitates with increasing vacancy concentration from plate-like to spherical shape.

Introduction

The precipitation of interstitial oxygen in Czochralski silicon is a topic which is of high interest since decades (see e. g. Ref. 1). Because of the complexity of the matter it is not yet fully understood. The importance of oxygen precipitation for the installation of a laterally homogeneous and effective internal getter for metallic impurities in silicon wafers requires the control of the parameters influencing oxide precipitate nucleation. Progress in the control and quantification of intrinsic point defects in silicon wafers by rapid thermal anneals (2) made it possible to investigate the impact of vacancies on oxide precipitate nucleation in detail (3,4).

The vacancy supersaturation installed by rapid thermal anneals impacts the getter efficiency of the oxide precipitates. On the one hand, vacancies built into the growing precipitates can reduce the lattice strain which is due to the different molecular volume of silicon and silicon oxide. On the other hand, it was reported that the getter efficiency of precipitates for chromium in RTA pre-treated wafers is decreased (5, 6).

In the last few decades, several criteria for efficient gettering of metallic impurities were published. Sueoka et al. determined a criterion which includes a critical length of plate-like oxide precipitates versus a critical density of precipitates for efficient intrinsic gettering of Ni (7). Seacrist et al. plotted the threshold for efficient gettering as a function both oxygen precipitate radius and density and confirmed Sueoka's criterion for spherical oxide precipitates (8). Hoelzl et al. showed that the getter efficiency depends on the total surface area of oxide precipitates in CZ-Si wafers and they defined a threshold value of the normalized inner surface as a criterion for efficient gettering of Cu and Ni (9). Kot et al. showed that pre-annealed wafers are in good agreement with Sueoka's threshold for efficient gettering, while data of RTA treated wafers are in good agreement with Hoelzl's criterion for efficient gettering (10).

Ex. 1016, Page 1023

An issue which is still of high interest is the elucidation of the role of vacancies for setting up an efficient getter for these metallic impurities. In this work, we investigate the getter efficiency of copper and nickel for different annealing conditions and vacancy supersaturations. Using the experimental results, we determined the threshold levels for efficient gettering. The results of these investigations are compared with the results of Hoelzl et al. (9).

Experimental

We used moderately boron doped (8-12 Ωcm) silicon wafers with an interstitial oxygen concentration of 6.4-6.5×10^{17} cm^{-3} (DIN 50438/1) and a diameter of 200 mm. These wafers received an RTA pre-treatment at 1250, 1225, and 1200 °C for 30 s in an Ar/O_2 atmosphere in order to establish certain vacancy concentrations in the wafers depending on the RTA temperature. Subsequently, the wafers were subjected to a nucleation anneal for 4 h and 8 h at a temperature in the range between 700 °C and 1000 °C. The BMD (bulk micro-defect) density was measured at the cleaved cross section for all samples after a growth anneal at 1000 °C for 16 h.

Pieces of the RTA pre-treated wafers were intentionally contaminated with Cu and Ni by scratching the metal as a dot on the backside after the nucleation anneal. The metal was diffused into the samples during a heat treatment at 900 °C for 10 min followed by fast cooling in air by simply pulling the samples out of the furnace. Finally, the samples were etched by Secco etchant (11) for 3 min. Using optical microscopy, the density of haze forming etch pits (12) was measured on the surface. The getter efficiency was determined by comparison of the density of haze forming etch pits on the annealed samples with the as-grown sample as a reference.

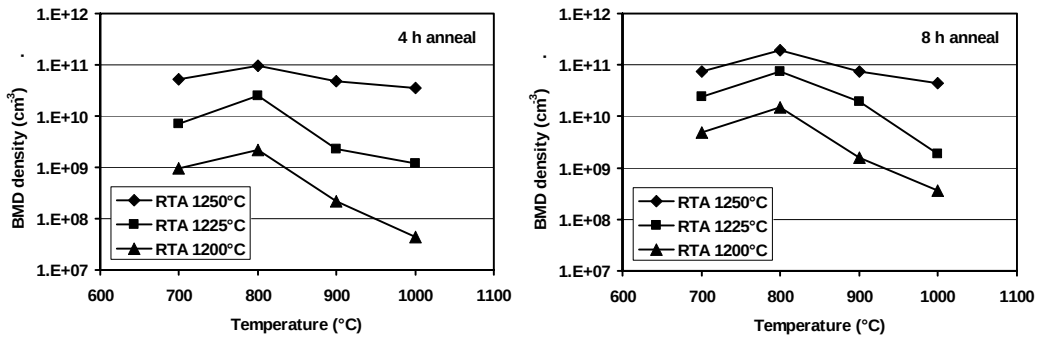


Figure 1. BMD density as a function of the temperature of the nucleation anneal for 4 h (left) and 8 h (right) shown for RTA pre-treatments at different temperatures for 30s. The initial concentration of interstitial oxygen was 6.4-6.5×10^{17} cm^{-3}.

Vacancy Dependent Oxide Precipitate Generation

Fig. 1 provides the results of the vacancy dependent oxide precipitate generation. The concentration of vacancies calculated according to (13) for RTA temperatures of 1200, 1225, and 1250 °C are 5.6×10^{12} cm^{-3}, 7.9×10^{12} cm^{-3}, and 1.1×10^{13} cm^{-3}, respectively. The concentrations obtained in this way are in a good agreement with vacancy concentrations measured by platinum diffusion. All curves possess a maximum at 800 °C which is in

agreement with the vacancy dependent nucleation rate (10). All peaks increase with increasing vacancy concentration what means with increasing RTA temperature. For high vacancy concentration, there are considerable BMD concentrations generated even at 1000 °C. Until now, this temperature was regarded as a pure growth temperature of oxide precipitates.

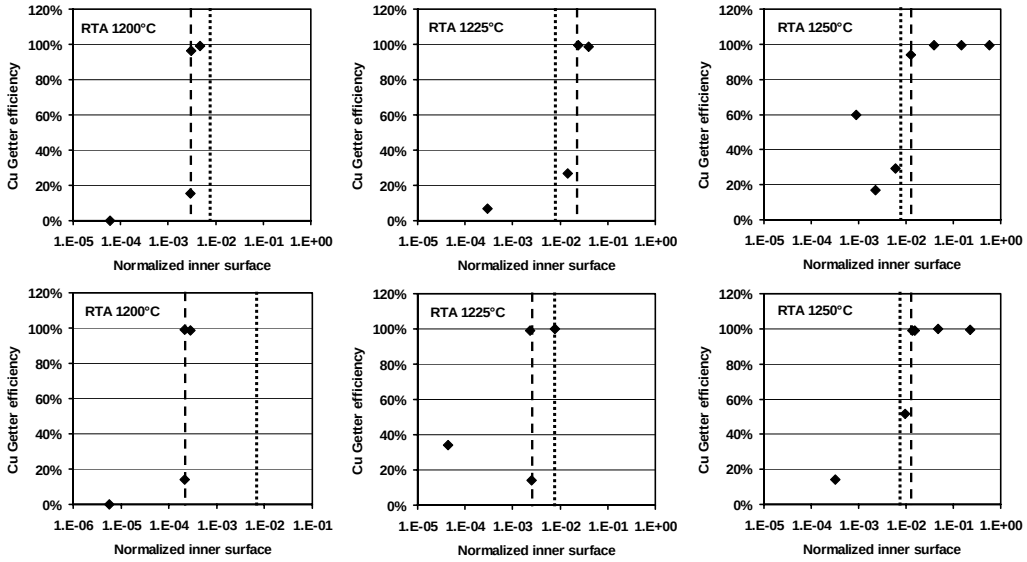


Figure 2. Getter efficiency for Cu as a function of the normalized inner surface shown for 4 h annealing (lower graphs) and 8 h annealing (upper graphs) and different temperatures of RTA pre-treatments. Please note the different scale of the bottom left graph.

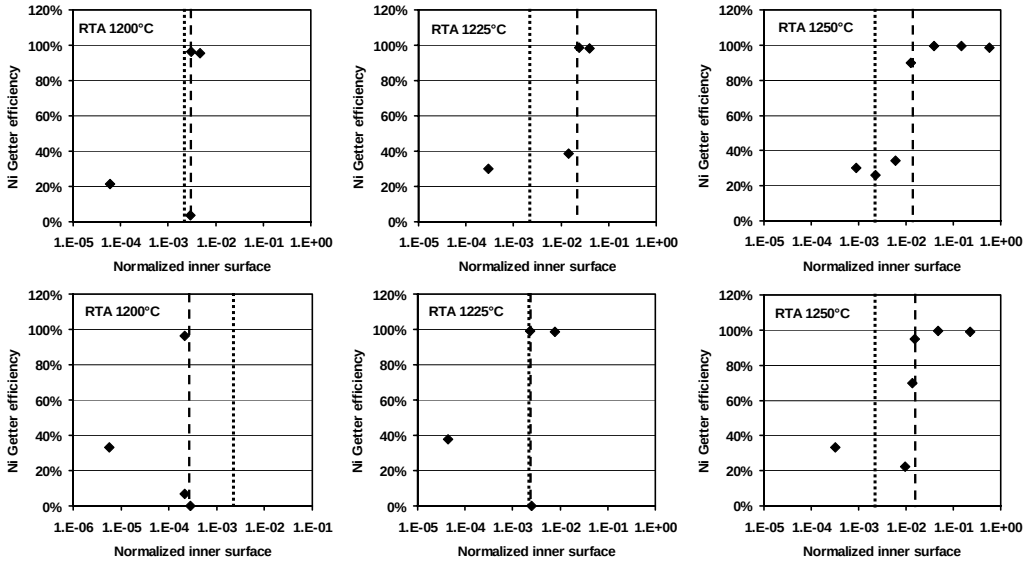


Figure 3. Getter efficiency for Ni as a function of the normalized inner surface shown for 4 h annealing (lower graphs) and 8 h annealing (upper graphs) and different temperatures of RTA pre-treatments. Please note the different scale of the bottom left graph.

Ex. 1016, Page 1025

The BMD curves show that the temperature region from 700-1000 °C would be suitable for the creation of enough efficient getter sites by annealing if the wafers would obtain an RTA pre-treatment in order to increase the vacancy concentration. Without the RTA pretreatment the resulting BMD density would be much lower and efficient gettering would require much larger oxide precipitates what means longer annealing times. In the next section, we describe the results of getter efficiency studies with RTA pre-treated wafers.

Getter Efficiency

Size and density of oxygen precipitates formed in silicon wafers during thermal treatment depend on time and temperature of the thermal treatment and on the concentration of interstitial oxygen c_{Oi} . For simplicity, we first assume that the precipitates are spherical. Then the normalized inner surface *S* according to Ref. (9) can be expressed as follows:

$$S = 4\pi \cdot r(t)^2 \cdot N_{BMD}(t) \cdot d_w \qquad [1]$$

with

$$r(t) = \left(2\frac{c_{Oi} - c_{Oi}^{eq}}{c_p} \right)^{0.5} \cdot \left(D_{Oi} \cdot t\right)^{0.5} \quad . \qquad [2]$$

In these equations, d_w is the thickness of the wafer, $N_{BMD}(t)$ is the density of bulk microdefects (BMDs) which are mainly oxygen precipitates, $r(t)$ the radius of oxygen precipitates after a certain time of annealing *t* according to Ref. (16), D_{Oi} is the diffusion coefficient of interstitial oxygen, c_p is the concentration of oxygen in the precipitate, and c_{Oi}^{eq} is the solubility of interstitial oxygen. In order to estimate critical values of *S* required for sufficient gettering, we plotted the getter efficiency for Cu and Ni against *S* what is shown in Figs. 2 and 3 for Cu and Ni, respectively. The point with the lowest value of *S* at which Cu or Ni are gettered by > 90 % is defined as the critical value of the normalized inner surface S_{crit} for efficient gettering. It is shown as a dashed line in the graphs. The value for S_{crit} from Hoelzl et al. is given as a dotted line for comparison. For both experiments with Cu and with Ni, it can be seen that the threshold value S_{crit} for efficient gettering is shifting to higher values with increasing RTA temperature. This means the higher the RTA temperature the larger is the surface area required for efficient internal gettering.

Discussion

First, we believed that the shift of S_{crit} is due to the incorporation of the excess vacancies, which increase in concentration with increasing RTA temperature, into the growing precipitates. This would decrease the elastic strain and thus the getter efficiency. However, this argument is weak because the excess vacancies diffuse out very fast and because their concentration is very low compared to the number of oxygen atoms in the precipitates.

Ex. 1016, Page 1026

Another possible explanation for the shift of S_{crit}, would be a morphological change of the precipitates. Plate-like precipitates would have a much larger surface area than spherical after the same growth period. In order to investigate this effect, the normalized inner surface was calculated using the diffusion limited growth of spheroidal oxide precipitates according to Vanhellemont (15). The aspect ratio between the major axis L and the minor axis d determines the shape of the oblate spheroids. In Eq. 1, the surface of the spheres was replaced by the surface of the oblate spheroids and S_{crit} was adjusted for the data of each graph in Figs. 2 and 3 to Hoelzl's threshold by varying the aspect ratio. The results are shown in Fig. 4. After the adjustment in the graphs on the left side of Fig. 4 all values are in agreement with Hoelzl's S_{crit}. From the corresponding aspect ratios given on the right side of Fig. 4, it can be concluded that for higher the RTA temperature the precipitates become increasingly spherical. This result is well in agreement with the results of Maddalon-Vinante et al. (5). They have found by FTIR measurements that during thermal treatment after a preceding RTA spherical instead of plate-like precipitates are formed which are not suitable to getter Cr. In our assumption, the getter efficiency of Cu and Ni per area remains constant but the BMD surface area for a given thermal budget changes due to a morphological change induced by the increasing RTA temperature. It can be also observed in Fig. 4 that the aspect ratio is increasing with the annealing time.

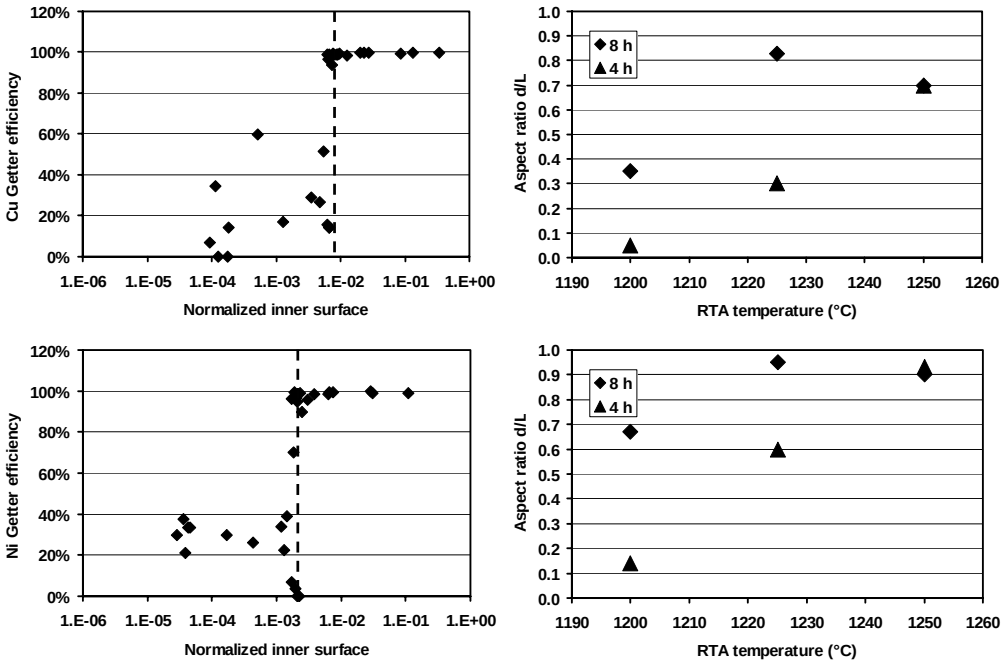


Figure 4. Getter efficiency for Cu (top left) and Ni (bottom left) as functions of the normalized inner surface based on oblate spheroids with different aspect ratios. The aspect ratios are shown as a function of the RTA temperature for Cu getter experiments (top right) and for Ni getter experiments (bottom right).

Conclusions

The influence of vacancy supersaturation installed by RTA pre-treatments in CZ silicon wafers on oxide precipitate nucleation was investigated in the temperature range 700-1000 °C. In summary, the results have demonstrated that precipitation is enhanced at 800 °C and increases with increasing vacancy concentration. Getter efficiency tests for Cu and Ni have shown that the threshold value of the normalized inner surface is shifting to higher values for increasing RTA temperature. This can be explained by a morphological change of the oxide precipitates with increasing vacancy concentration from plate-like to spherical shape.

Acknowledgments

The authors thank M. Boy from Siltronic AG for the RTA pre-treatments.

References

1. A. Borghesi, B. Pivac, A. Sassella, and A. Stella, *J. Appl. Phys.* **77**, 4169 (1995).
2. R. Falster, M. Pagani, D. Gambaro, M. Cornara, M. Olmo, G. Ferrero, P. Pichler, and M. Jacob, *Solid State Phenom*., **57-58**, 129 (1997).
3. G. Kissinger, D. Kot, J. Dabrowski, V. Akhmetov, A. Sattler, and W. von Ammon, in *High Purity Silicon X*, C. Claeys, R. Falster, M. Watanabe, and P. Stallhofer, Editors, in: ECS Transactions Vol. 16 No. 8, 97 (2008).
4. R. Falster, V. V. Voronkov, and F. Quast, *phys. stat. sol. (b)*, **222**, 219 (2000).
5. C. Maddalon-Vinante, E. Ehret, and D. Barbier, *J. Appl. Phys.* **79**, 2707 (1996).
6. C. Maddalon-Vinante, J. P. Vallard, and D. Barbier, *J. Electrochem. Soc.* **142**, 560 (1995).
7. K. Sueoka, M. Akatsuka, T. Ono, E. Asayama, Y. Koike, N. Adachi, S. Sadamitsu, and H. Katahama, in *High Purity Silicon VI*, C. Claeys, P. Rai-Choudhury, M. Watanabe, P. Stallhofer, and H. J. Dawson, Editors, PV 2000-17, p. 164, The Electrochemical Society Proceedings Series, Pennington, NJ (2000).
8. M. Seacrist, M. Stinson, J. Libbert, R. Standley and J. Binns, in: *Semiconductor Silicon 2002*, H. R. Huff, L. Fabry, and S. Kishino, Editors, PV 2002-2, p. 638, The Electrochemical Society Proceedings Series, Pennington, NJ (2002).
9. R. Hoelzl, M. Blietz, L. Fabry and R. Schmolke, in *Semiconductor Silicon 2002*, H. R. Huff, L. Fabry and S. Kishino, Editors, PV 2002-2, p.608, The Electrochemical Society Proceedings Series, Pennington, NJ (2002).
10. D. Kot, G. Kissinger, W. Häckl, A. Sattler, and W. von Ammon, in *High Purity Silicon X*, C. Claeys, R. Falster, M. Watanabe, and P. Stallhofer, Editors, ECS Transactions Vol. 16 No. 8, 207 (2008).
11. F. Secco d'Aragona, *J. Electrochem. Soc.*, **199**, 948 (1972).
12. K. Graff, *Metal Impurities in Silicon–Device Fabrication*, Springer-Verlag, Berlin, Heidelberg, New York (1995).
13. G. Kissinger, J. Dabrowski, A. Sattler, C. Seuring, T. Müller, H. Richter, and W. von Ammon, *J. Electrochem. Soc.*, **154**, H454 (2007).
14. J. Vanhellemont, *J. Appl. Phys.*, **78**, 4297 (1995).

ECS Transactions, 18 (1) 1001-1011 (2009)
10.1149/1.3096564 © The Electrochemical Society

Oxygen Precipitation in Heavily Phosphorus-doped Czochralski Silicon

Yuheng Zeng, Deren Yang, Xiangyang Ma, Xinpeng Zhang, Lixia Lin, Duanlin Que

State Key Laboratory of Silicon Materials and Department of Materials Science and Engineering, Zhejiang University, Hangzhou 310027, People's Republic of China

Oxygen precipitation in heavily phosphorous (P)-doped Czochralski silicon (CZ-Si) subjected to the ramping or two-step anneals was investigated. It was revealed that the grown-in oxygen precipitates have significant effect on oxygen precipitation behaviors during the subsequent annealing, and then resulted in distinctive oxygen precipitation for the wafers form different positions of the ingot. Moreover, it was found that both the seed and tang-end wafers processed significantly intense nucleation at 650 ^{o}C, which was related to phosphorus related precipitates. It was pointed out that with the increase of the annealing temperature, the nucleation rate of oxygen precipitation due to the enhancement of heavily P-doping would gradually weaken but the one due to oxygen clustering became pronounced. Thus, it was understandable that the tang-end wafer annealed at around 750 ^{o}C processed higher density of oxygen precipitates than the seed-end wafer, while it annealed at 1000 ^{o}C processed a lower one.

1. Introduction

Epitaxial silicon wafer has prevalently been used for ultra-large integrated circuit (IC) in order to increase device yield, because it processes low grown-in crystal defects. Heavily doped Czochralski silicon (CZ-Si) is considered as one of the most important ideal substrates for epitaxial silicon wafer. Since heavily doped CZ-Si processes low resistivity and then suppresses latch-up effect and reduces power loss, it is of importance for advanced CMOS devices or power devices [1]. Usually, internal gettering essentially determined by oxygen precipitation in substrates was required to increase the production yield during the fabrication of IC devices [2, 3]. Through oxygen precipitation in CZ-Si has been intensively investigated in the past decades, most of the researches are concerned with lightly doped CZ-Si, and the investigation for heavily doped CZ-Si is insufficient [4, 5]. As we know, heavily phosphorus (P) doped CZ-Si is considered as a promising substrate due to the lowest resistivity among n-type heavily doped CZ-Si [6, 7];

Ex. 1016, Page 1029

however, to our knowledge, its oxygen precipitation behavior has been hardly reported.Thus, the investigation for oxygen behaviors in heavily P-doped CZ-Si is practically necessary.

In our previous work, it was found that heavily P-doped CZ silicon possess high density of grown-in oxygen precipitates those would grow up during one-step high temperature annealing [8]. Moreover, it was found that the performances of oxygen precipitation in the wafers coming from various sections of an ingot were significantly different, which became an interest matter worthy to reveal.

In this presentation, the oxygen precipitation behaviors from seed and tang-end heavily P-doped CZ-Si subjected to the ramping or two-step anneals were investigated. It was found that as-received heavily P-doped CZ-Si had high densities of grown-in oxygen precipitates, which could result in the retardant effect of oxygen precipitation for the seed-end wafer. Further investigation revealed that 650 ^{o}C annealing led to significantly oxygen precipitation, which was considered as the reason for the high density of grown-in oxygen precipitates in heavily P-doped CZ silicon. Finally, the mechanism for intense oxygen precipitation nucleation at 650 ^{o}C was discussed.

2. Experiments

A heavily P-doped <100>-oriented CZ-Si ingot was employed in this experiment. The slices taken from the seed and tang-end of the ingot were processed into one-side mirror-polished wafers and then used as samples. The resistivity of the seed and tang-end samples were about 0.0015 and 0.001 Ω.cm, respectively. The interstitial oxygen concentrations ([O]s) were measured as 10.5×10^{17} and 9.5×10^{17} cm^{-3} in the seed and tang-end samples, while the phosphorus concentrations ([P]s) were about 4.0×10^{19} and 5.6×10^{19} cm^{-3} in the seed and tang-end samples. Both the [O]s and [P]s were determined by a Cameca 4f secondary ion mass spectroscope (SIMS).

The samples sized in 2×3 cm^2 taken from the seed and tang-end CZ-Si wafers were, respectively, divided into two groups. The first group was the as-received samples, they received the anneal of 1050 ^{o}C/16 h ramped up from 650, 750, 850 and 950 ^{o}C with a rate of 1 ^{o}C·min^{-1}, respectively. The second group received the prior treatment of 1230 ^{o}C/20 min conventional furnace annealing (CFA) to substantially dissolve grown-in oxygen precipitates, and then were

withdrawn slowly form the furnace in order not to induce slips in the samples. Subsequently, they received the following anneals schemes: (1) (650, 750, 850 and 950) ^{o}C/4 h + 1050 ^{o}C/16 h with a ramping rate of 1 ^{o}C·min^{-1}, respectively; (2) 1000 ^{o}C/ (16, 48) h. It should be mentioned that the above heat treatments were carried out under argon (Ar) atmosphere. After anneal, the samples were cleaved and preferentially etched in Y1 etchant [CrO_3 (0.5 mol/L):HF(49%) = 1:1] for 10 min [9], and the cross-sectional of bulk micro-defects (BMDs) were observed by an Olympus MX50 optical microscope (OM).

3. Results and discussions

Fig. 1 shows the cross-sectional OM micrographs of BMDs in the as-received seed and tang-end heavily P-doped CZ-Si subjected to the anneal of 1050 ^{o}C for 16 h ramped up from 650, 750, 850 and 950 ^{o}C with a rate of 1 ^{o}C·min^{-1}, respectively. As can be seen, the BMD density of the seed-end samples increased with the increase of the the starting temperature of ramping anneal, indicating that the more oxygen precipitate nuclei could grow up. Contrarily, the BMD density of the tang-end CZ-Si decreased with the increase of the starting temperature of ramping anneal. In the cases of 850 and 950 ^{o}C starting temperature, the BMD densities of the seed and tang-end samples became comparable. Therefore, it is concluded that the behaviors of oxygen precipitation from various positions of a heavily P-doped ingot were significant different during ramping anneal, especially in the cases that the starting temperature of ramping anneal was lower than 750 ^{o}C.

It is well known that oxygen precipitates whose sizes larger than the critical size at annealing temperatures can grow up, whereas the ones smaller than the critical size would shrink during annealing [10]. The density of oxygen precipitates decreases with the sizes increases according to the density-size profile of oxygen precipitates [10]. Consequently, it is believed that the density of oxygen precipitates generated during ramping annealing suggested the profile of grown-in oxygen precipitates, and then lower starting ramp-up temperatures results in more oxygen precipitates [10, 11]. Moreover, generally speaking, due to the longer thermal history during crystal growth, the seed-end wafer of a CZ-Si ingot processes more density of oxygen precipitates than the tang-end wafer. Obviously, from the result of Fig. 1, it was found that the oxygen precipitation behaviors in the as-received seed-end heavily P-doped CZ-Si did not agree with the commonsense mentioned

above.

Fig. 2 shows the BMD densities in the CFA pre-treated seed and tang-end CZ-Si subjected to the anneal of (650, 750, 850 and 950) ^{o}C/4 h + 1050 ^{o}C/16 h with a ramping rate of 1 ^{o}C·min^{-1}. As can be seen, the oxygen precipitation behaviors from the seed and tang-end sample became almost coincident during the anneal, meaning that the higher starting ramp-up temperature the less oxygen precipitates were generated. Different form the as-received samples, the higher temperature CFA pre-treatment has dissolved grown-in oxygen precipitates so that the influence of thermal history of the wafers during crystal growth was removed. This result indicates that it is t no essential differences of oxygen precipitation during the annealing between the seed-end and tang-end samples. The abnormal oxygen precipitation behaviors in the seed-end sample shown in Fig. 1 were probably resulted from the thermal history, i.e. grown-in oxygen precipitates. Furthermore, from Fig. 2, it was found that in both kind of the samples, the annealing at 650 ^{o}C led to significantly high density of BMDs while those at 750 and 850 ^{o}C resulted in much lower densities, suggesting the stronger nucleation rate of oxygen precipitation at 650 ^{o}C but a noticeably weaker rate at higher temperatures. It is due to the super-saturation of oxygen in CZ Si during annealing in low temperature range (600 to 800 ^{o}C) [4].

It was also found in Fig. 2 that the BMD density in the tang-end wafer annealed at 750 ^{o}C was higher than that in the seed-end wafer, indicating that the tang-end wafer could produce more nuclei than the seed-end wafer. The results can be further confirmed by Fig. 3 which are the cross-sectional OM micrographs of BMDs in the CFA pre-treated seed and tang-end heavily P-doped CZ-Si subjected to the subsequent anneal of (650, 750 and 850) ^{o} C/4 h + 1050 ^{o}C/16 h with a ramping rate of 1 ^{o}C·min^{-1}. Since the P concentration in the tang-end samples was higher than that in the seed-end ones, it was considered that the nucleation at 750 ^{o}C was remarkably influenced by the P concentration. Regarding to annealing at 850 ^{o}C and above, the density of oxygen precipitates was nearly the same in both the seed and tang-end wafers, indicating that the heavily P-doping did not have significantly effect on oxygen precipitation any more at higher temepratures.

It has well demonstrated that silicon phosphide precipitates can be present in CZ-Si after thermal diffusion processes of typical device technology [12-15]. For instances, A.Armigliato et al. [12, 13] reported the occurrence of SiP precipitates in CZ-Si after the phosphorous predeposition

at 920 ^{o}C and 1000 ^{o}C using $POCl_3$. S.Solmi et al. [15, 16] also reported the formation of largely monoclinic and partially orthorhombic SiP precipitates in the silicon wafers subjected to annealing at 800-1000 ^{o}C, following the implantation of P with a high fluence of 1.5×10^{17} cm^{-2}. The above-mentioned work has clearly demonstrated that the formation of silicon phosphide precipitates in silicon is driven by the super-saturation of P. Obviously, when P concentration is over super-saturated, P atoms can not be present as solute one in silicon lattice anymore; instead it constituted silicon phosphide precipitates. The maximum annealing temperature for the formation of silicon phosphide precipitates increases with P concentration decreasing. Actually, in our other report, we demonstrated that the nucleation in heavily P-doped CZ-Si at 650 ^{o}C was resulted from silicon phosphide precipitation [17]. Herein, it was also considered that the nucleation of oxygen precipitates at 650 ^{o}C was enhanced by heavily P-doping, and then the enhancement effect gradually weakened with the increase of annealing temperatures.

Moreover, in view of the fact that intense oxygen precipitate nuclei were formed at 650 ^{o}C and the fact that a heat preservation of tens of hours in the temperature range of 200-650 ^{o}C [18] during silicon crystal growth, it was reasonably inferred that a quantities of grown-in oxygen precipitates was formed during the crystal pulling process, which was also confirmed by our previous report [8]. Therefore, it was considered that the oxygen precipitates in the as-received seed-end wafer were retarded by extremely high density of grown-in oxygen precipitates during the ramping anneal.

The nucleation rate at 750 ^{o}C and above was significant weakened; however, the mechanism was not yet clear. It was only known that with temperature increasing, the nucleation based on silicon phosphide precipitates would be weak; instead, the one based on oxygen clusters would be gradually dominative. From the remarkable differences of the precipitation behavior at 750 ^{o}C for the seed end and tang end wafers, it was inferred that the nucleation at 750 ^{o}C and above was gradually dominative by oxygen clusters. However, even in this case, it is hardly said that the oxygen precipitate nucleation at 750 ^{o}C was definitively irrelevant with the phosphorus related defects. Form Fig. 3, it can be seen that the heavily P-doping effect still exerted remarkable influences on oxygen precipitation at 750 ^{o}C. Overall, the mechanism of oxygen precipitate nucleation in heavily P-doped CZ-Si required further investigations.

To reveal the oxygen effect on oxygen precipitation in heavily P-doped CZ Si, a prolong

Ex. 1016, Page 1033

annealing at 1000 ^{o}C following 1230 ^{o}C CFA pre-treatment was carried out for the seed and tang-end CZ-Si, shown in Fig. 4. It can be seen in the figure that no oxygen precipitates were present after 16 h annealing in both of the seed and tang-end CZ-Si, indicating that the oxygen precipitates nucleation was weak at 1000 ^{o}C. With extending of the annealing time to 64 h, the density of oxygen precipitates in the seed-end sample was higher than that in the tang-end wafer.Therefore, it was considered that at 1000 ^{o}C, the heavily P-doping probably has no longer effect on oxygen precipitation which was basically ascribed to the super-saturation of oxygen. Therefore, the seed-end wafer would led to more oxygen precipitates in the extending anneal because of its higher oxygen concentration.

4. **Conclusions**

The oxygen precipitation in heavily P-doped CZ-Si subjected to the ramping or two-step anneal was investigated. It was revealed that the oxygen precipitation from the as-received seed and tang-end wafers during ramping anneal were distinctive. It was found that the oxygen precipitation in the as-received seed-end wafer was retarded by the higher density of grown-in oxygen precipitates; while the retardant effect did not occur in the tang-end wafer. After eliminating the grown-in oxygen precipitates, both of the seed and tang-end wafers processed significantly higher nucleation rate of oxygen precipitates at about 650 ^{o}C, which was ascribed to the fact that the heavily P-doping introduced phosphorus related precipitates and enhanced the oxygen precipitation. Moreover, it was believed that the high density of grown-in oxygen precipitates was resulted from heavily P-doping. When the increase of annealing temperatures, the nucleation rate of oxygen precipitates due to heavily P-doping would be gradually weak but the one due to oxygen clusters became pronounced. It was understandable that the tang-end wafer annealed at around 750 ^{o}C processed higher density of oxygen precipitates than the seed-end wafer, while it annealed at 1000 ^{o}C had a lower one.

Acknowledgments

Ex. 1016, Page 1034

The authors would like to thank the financial supports for Natural Science Foundation of China (No. 50832006, 60876001 and 50672085), the Research Fund for the Doctoral Program of Higher Education (No. 200803350043), Program PCSIT, Program 973 (No. 2007CB613403), Zhejiang provincial and Ningbo municipal science and technology program.

References

1. R. R. Troutman, G. D. Ji and W. H. Lu, Latch-up in CMOS technology: the problem and its cure, Kluwer Academic, Boston, 1986.
2. R. Falster and V. V. Voronkov, *Mat. Sci. Eng. B-Solid.*, **73**, 87 (2000)
3. M. L. Polignano, G. F. Cerofolini, H. Bender and C. Claeys, *J. Appl. Phys.*, **64**, 869 (1988)
4. A. Borghesi, B. Pivac, A. Sassella and A. Stella, *J. Appl. Phys.*, **77**, 4169 (1995)
5. F. Shimura, Oxygen in Silicon, Academic Press, New York, 1994.
6. H. Kodera, *Jpn. J. Appl. Phys.*, **2**, 527 (1963)
7. H.-D. Chiou, *J. Electrochem. Soc.*, **147**, 345 (2000)
8. Y. Zeng, D. Yang, X. Ma, J. Chen and D. Que, *Mat. Sci. Eng. B-Solid.*, **accepted**,
9. K. H. Yang, *J. Electrochem. Soc.*, **131**, 1140 (1984)
10. N. Inoue, J. Osaka and K. Wada, *J. Electrochem. Soc.*, **129**, 2780 (1982)
11. L. Gong, X. Ma, D. Tian, L. Fu and D. Yang, *Semicond. Sci. Technol.*, **20**, 1 (2005)
12. A. Armigliato, D. Nobili, M. Servidori and S. Solmi, *J. Appl. Phys.*, **47**, 5489 (1976)
13. A. Armigliato, M. Servidori, S. Solmi and I. Vecchi, *J. Appl. Phys.*, **48**, 1806 (1977)
14. D. Nobili, A. Armigliato, M. Finnetti and S. Solmi, *J. Appl. Phys.*, **53**, 1484 (1982)
15. S. Solmi, A. Parisini, R. Angelucci, A. Armigliato, D. Nobili and L. Moro, *Phys. Rev. B*, **53**, 7836 (1996)
16. S. Solmi and D. Nobili, *J. Appl. Phys.*, **83**, 2484 (1998)
17. Y. Zeng, X. Ma, W. Wang, D. Tian, L. Gong and D. Yang, *J. Appl. Phys.*, **Summitted**,
18. A. Virzi and M. Porrini, *Mat. Sci. Eng. B-Solid.*, **17**, 196 (1993)

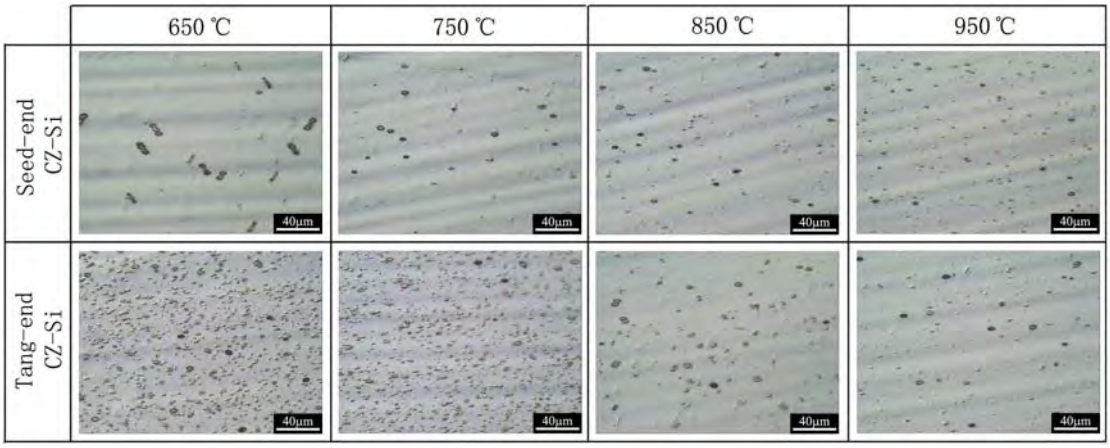


Fig. 1. Cross-sectional OM micrographs of BMDs in the as-received seed and tang-end heavily P-doped CZ-Si subjected to the anneal of 1050 ^{o}C for 16 h ramped up from 650, 750, 850 and 950 ^{o}C with a rate of 1 ^{o}C·min^{-1}, respectively.

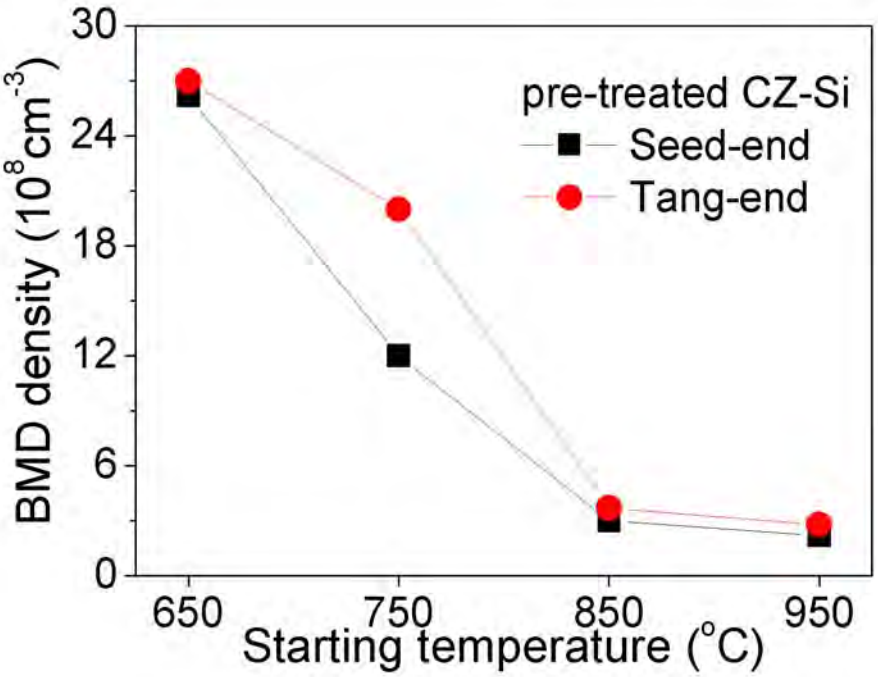


Fig. 2. BMD densities in the CFA pre-treated seed and tang-end CZ-Si subjected to the anneal of (650, 750, 850 and 950) ^{o}C/4 h + 1050 ^{o}C/16 h with a ramping rate of 1 ^{o}C·min^{-1}.

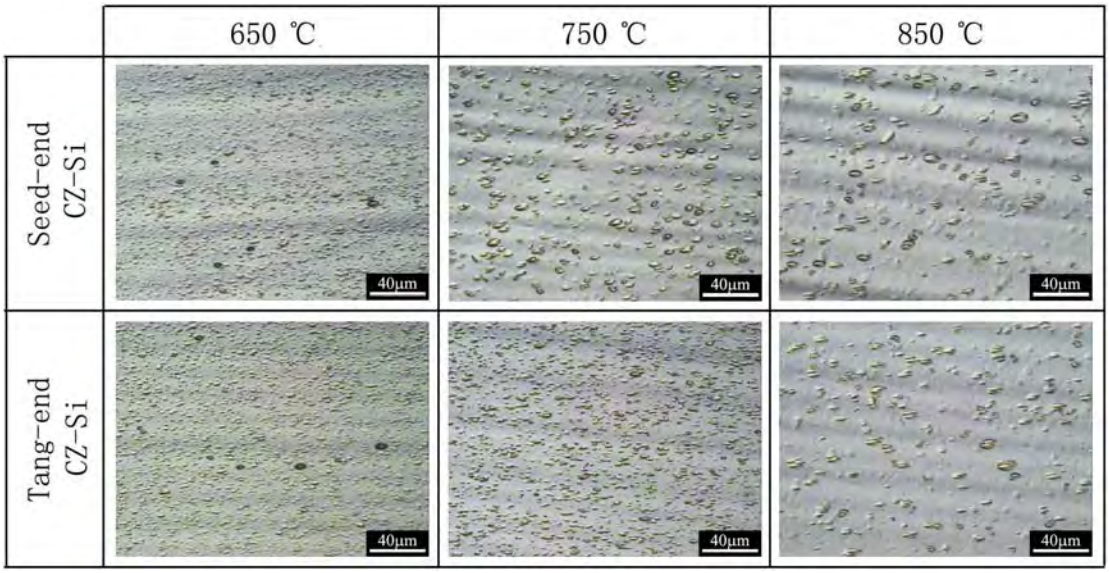


Fig. 3. Cross-sectional OM micrographs of BMDs in the CFA pre-treated seed and tang-end heavily P-doped CZ-Si subjected to the anneals of (650, 750 and 850) ^{o}C/4 h + 1050 ^{o}C/16 h with a ramping rate of 1 ^{o}C·min^{-1}.

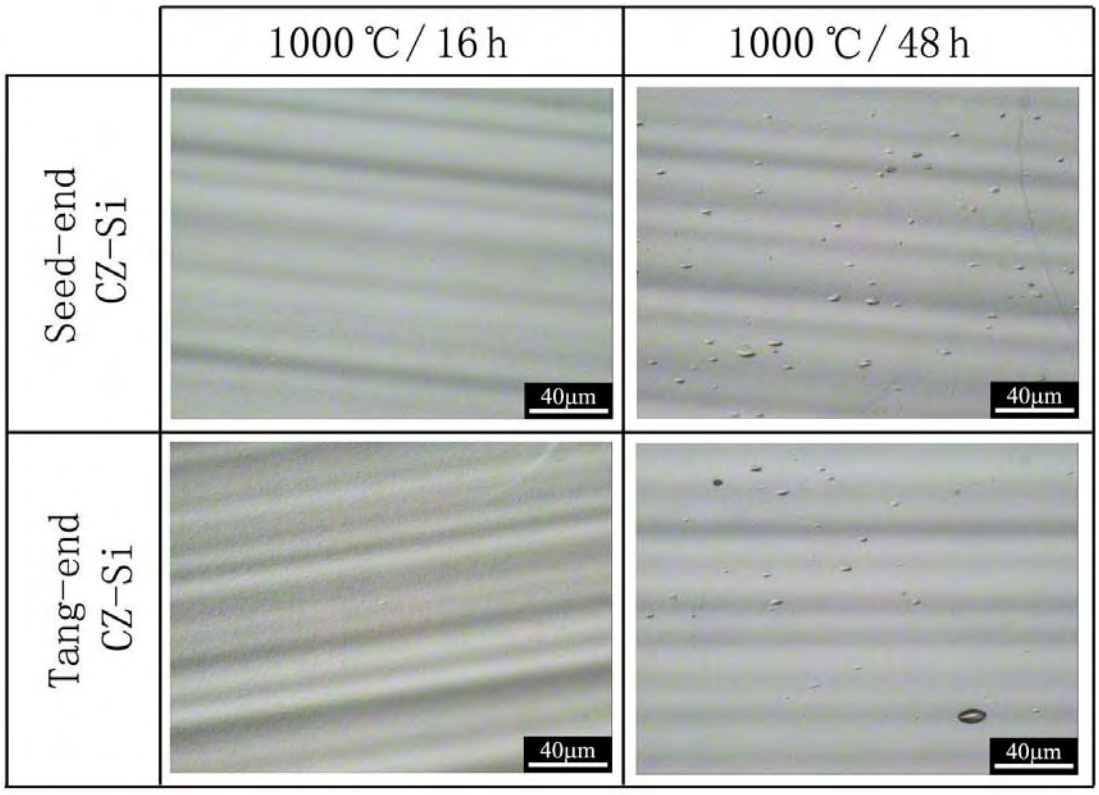


Fig. 4. Cross-sectional OM micrographs of BMDs in the CFA pre-treated seed and tang-end heavily P-doped CZ-Si subjected to the anneals of 1000 ^{o}C/ (16 and 48) h.

ECS Transactions, 18 (1) 1013-1018 (2009)
10.1149/1.3096565 © The Electrochemical Society

Interactions of Cu and Ni impurities with vacancy-related point defects in Czochralski-grown Si crystals

V. P. Markevich[a], A. R. Peaker[a], I. F. Medvedeva[b], V. E. Gusakov[b], L. I. Murin[b], and B. G. Svensson[c]

[a] Scholl of Electrical and Electronic Engineering, University of Manchester, Manchester M60 1QD, United Kingdom
[b] Scientific-Practical Materials Research Centre of NAS of Belarus, Minsk 220072, Belarus
[c] Department of Physics, University of Oslo, 0316 Oslo, Norway

It has been inferred from DLTS measurements on electron-irradiated Si samples contaminated with copper and nickel that divacancies (V_2) and vacancy-oxygen complexes (VO) are effective traps for mobile Cu and Ni interstitial atoms. Interactions of these impurities with VO and V_2 result in the enhanced annealing out of the VO and V_2 defects as well as in the formation of new electrically active centers. An enhanced annealing of the VO and V_2 defects in Si:Cu crystals is accompanied by an appearance of electron traps with the energy levels at E_c – 0.60 eV and E_c – 0.17 eV. It is argued that these levels are related to acceptor levels of Cu-VO and Cu-V_2 complexes. The stable configuration of the Cu-VO complex has been determined by means of quantum-chemical calculations. In Si:Ni crystals an energy level at E_c – 0.365 eV is related to a complex incorporating at least one Ni atom and a radiation-induced defect.

Introduction

Transition-metal contaminants are ubiquitous in Si materials and devices, being introduced both during wafer growth and in subsequent processing (1, 2). Cu and Ni are especially troublesome for technology of Si-based devices. Cu and Ni are among the fastest diffusers in Si and can introduce into the material a variety of electrically active defects, some of them being detrimental to the performance of devices (1, 2).

Copper at low concentrations in crystalline silicon exists predominantly as the Cu_i^+ interstitial. As copper is now used for interconnects on extremely scaled integrated circuits, the possibility of Cu contamination during technology processes has increased in recent years. In consequence there is strong interest in understanding the details of interactions between Cu and common defects and impurities in silicon particularly those resulting from implantation or plasma damage. As the solubility of copper in Si is not high at moderate temperatures but it is a fast diffuser it tends to precipitate near vacancy-related defects, stacking faults, grain boundaries, oxygen precipitates, or out diffuse to the surface of samples (1-3). Some copper related defects are electrically and optically active and can strongly affect generation-recombination processes.

Available information on interactions of Cu atoms with radiation-induced defects in Si crystals and on the properties of resulting complexes is rather limited. Theoretical calculations predict Cu interactions with vacancy-related defects {vacancy (V), divacancy (V_2), vacancy-oxygen pair (VO), hexa-vacancy (V_6)} as well as with a Si self-

interstitial (3, 4). Minimum energy configurations of the resulting complexes have been found in Ref. (3), however, no information on the electronic properties of these complexes has been reported. There are only some data on the properties of a Cu atom at a substitutional site (Cu_s), which results from the reaction $Cu_i + V \rightarrow Cu_s$. Cu_s in Si has been identified by deep-level transient spectroscopy (DLTS) (5, 6). It has three gap levels at $E_v + 0.20$ eV (0/+), $E_v + 0.41$ eV (-/0), and $E_c - 0.17$ eV (- -/-).

Results of a DLTS study of the interaction between copper and irradiation-induced defects in silicon were reported in Refs. (7, 8). The results were obtained with the use of Schottky-barrier structures formed by the deposition of copper on *n*- and *p*-type silicon irradiated with 2-MeV electrons at room temperature. It was found that the annealing rate of the divacancy is anomalously high at temperatures below 150 ºC in both *n*- and *p*-type Si. This was attributed to passivation of the V_2 centres by Cu atoms through the formation of neutral complexes. Further, a level at $E_v + 0.52$ eV was assigned to a complex formed by a reaction between Cu and carbon-oxygen centres (C_iO_i) which give rise to the "ordinary" irradiation-induced level at $E_v + 0.35$ eV [7]. This latter level, which is normally stable up to temperatures of ~350 ºC, annealed out rapidly at 100 ºC with the simultaneous formation of a Cu-associated level at $E_v + 0.52$ eV.

There is no reliable information on the interaction of Ni with radiation-induced defects. It can be suggested that interstitial Ni atoms interact with vacancies. This reaction results in the appearance of Ni atoms at substitutional sites. It was speculated that gap levels at $E_v + 0.18$ eV and $E_c - 0.47$ eV might be related to Ni_s (1).

Results of a study of the influence of Cu and Ni contaminations on radiation-induced defect reactions in phosphorus-doped Czochralski-grown silicon (Cz-Si) crystals are reported in the present work. The results have been obtained by means of the Hall effect, and DLTS techniques with supporting theoretical modeling of defects.

Experimental Details

The samples used in this study were cut from a phosphorus doped {$N_P = (7\text{-}8)\times10^{14}$ cm^{-3}} Cz-Si crystal. The concentrations of interstitial oxygen and substitutional carbon atoms in the crystal were about 1×10^{18} cm^{-3} and 5×10^{15} cm^{-3}, respectively. Copper and nickel were introduced into the samples from metallic layers on the surface by in-diffusion heat-treatments at 600 and 650 °C for Cu and Ni, respectively. The in-diffusion heat-treatments were followed by fast cooling. According to the literature data (9), such heat-treatments can result in the introduction of up to $(1.0\text{-}1.5)\times10^{15}$ cm^{-3} Cu and Ni atoms into the Si samples. A few samples, which did not have a metallic layer on the surface, were subjected to the analogous heat-treatments at 600 °C. After the heat-treatments a surface layer with the thickness of about 100 μm has been removed from the samples by mechanical polishing and etching in a CP4 acid mixture. The Cu(Ni)-doped samples were irradiated with 4 MeV electrons at room temperature together with the as-grown and heat-treated samples, which were not contaminated with Cu(Ni). For all the samples Schottky diodes were fabricated by thermal evaporation of Au on surfaces etched in a $1HF+5HNO_3$ acid mixture. Deep electronic levels were characterized with conventional DLTS and Laplace DLTS techniques (10). Concentrations of free electrons in the samples were measured with the use of a conventional Hall effect technique in the temperature range 77-400 K. 30-minute isochronal anneals of the irradiated and non-irradiated samples were carried out in a dry nitrogen ambient in the temperature range 75-400 °C with 25 °C increments.

Experimental Results

DLTS measurements on the samples, which were subjected to heat-treatments at 600 and 650 °C with and without in-diffusion of Cu(Ni), have shown that these treatments have not resulted in the introduction of deep level traps with concentrations higher than 10^{11} cm^{-3} into the upper half of the Si band gap. Only a very weak broad signal with a maximum at about 210 K for a rate window of 80 s^{-1} has been observed in the DLTS spectra of the Cu-doped samples. The heat-treatments caused no detectable changes in the temperature dependencies of free electrons {*n(T)*} for the samples. Further, 30-minute isochronal anneals of the heat-treated and as-grow n samples in the temperature range 75-400 °C did not result in any significant changes in the DLTS spectra as well as in the *n(T)* dependencies.

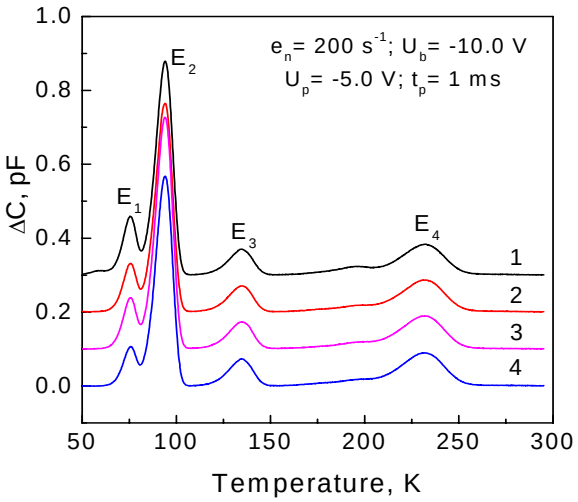


Figure 1. DLTS spectra of Cz-Si samples irradiated with 4 MeV electrons at room temperature: (1) as-grown, (2) and (3) heat-treated at 600°C for 60 min without and with a Cu layer on the surface and (4) heat-treated at 650°C for 180 min with a Ni layer on the surface. The fluence of irradiation was 1×10^{15} cm^{-2}. The spectra have been recorded with the measurement settings, which are shown in the figure. The spectra (1) to (3) are shifted on vertical axis for clarity.

Fig. 1 shows the conventional DLTS spectra of the Cu- and Ni-doped, as-grown and heat-treated samples after an irradiation with 4 MeV electrons. It should be noted that the spectra have been recorded with a reverse bias voltage of –10.0 V and a pulse voltage of –5.0 V. Such measurement settings have been chosen in order to move the region probed by DLTS farther from the surface and, thus, to reduce possible effects of hydrogen on radiation-induced defects reactions. Hydrogen is introduced into the sub-surface layers of Si samples during wet chemical etching and interacts effectively with radiation-induced defects (11). Four distinct peaks, labeled as E_1 to E_4 in Fig. 1, with maxima at about 75, 95, 135 and 230 K for an emission rate window of 200 s^{-1} are observed in the spectra of all the samples. A Laplace DLTS analysis has shown that the activation energies for electron emission (E_e) from the dominant traps are the same in all the samples and the E_e values are 0.13, 0.165, 0.24 and 0.425 eV for the E_1 to E_4 traps, respectively. The DLTS spectra shown in Fig.1 resemble those reported in the literature for Cz-Si crystals irradiated with MeV electrons at room temperature (11-14). On the basis of a comparison of the measured trap parameters (activation energies for electron emission and values of electron capture cross section) with the known values for defects induced by electron

irradiation in Cz-Si crystals (11-14), we assign the E_1 peak to an acceptor level of the self-interstitial–dioxygen complex (IO_2), the E_2 peak to an acceptor level of the vacancy–oxygen (VO or A centre) complex, the E_3 peak to the second acceptor level of the divacancy and the E_4 peak to the first acceptor level of the divacancy with an admixture of another level with a low concentration. This minor level disappears completely together with a low-temperature shoulder of the E_4 peak after annealing at 100 °C.

It has been found that the heat-treatments at 600 °C with and without in-diffusion of Cu do not influence significantly the annealing behavior of the IO_2 complex. This defect has disappeared completely in all the samples after annealing at 125 °C. However, the high-temperature treatments affect the annealing behavior of two vacancy-related defects, the A centre and the divacancy. The changes in concentrations of the vacancy-related defects in the irradiated samples upon isochronal annealing are presented in Fig. 2.

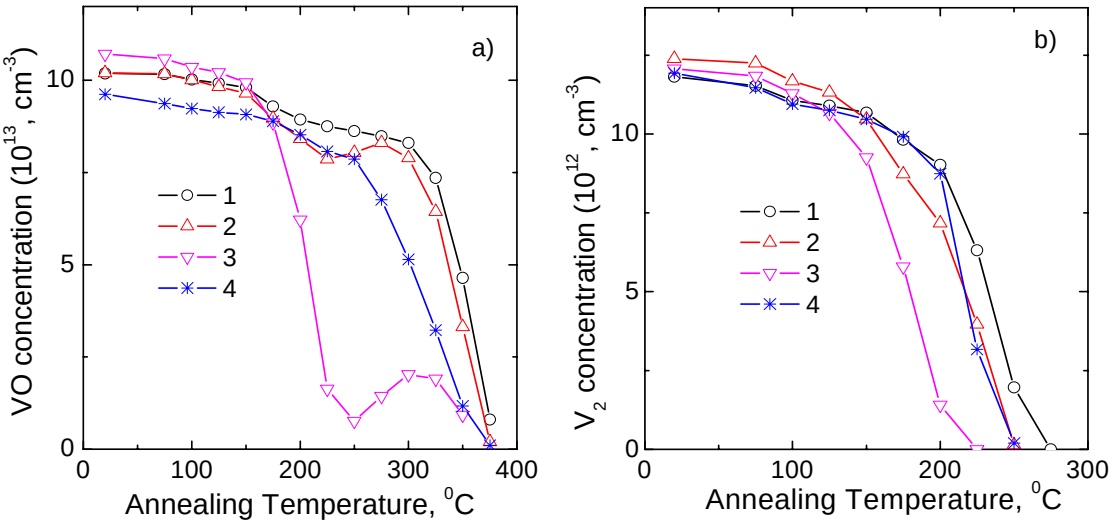


Figure 2. Changes in concentrations of a) vacancy-oxygen complex and b) divacancy upon 30-min isochronal annealing of electron-irradiated Cz-Si samples: (1) as-grown, (2) and (3) heat-treated at 600°C for 60 min without and with a Cu layer on the surface and (4) heat-treated at 650°C for 180 min with a Ni layer on the surface.

Fig. 2a shows that the A centre disappears in the Cu- and Ni-doped sample upon annealing at significantly lower temperatures compared to the temperature of the VO elimination in the uncontaminated samples (300-375 °C). This effect is especially pronounced in Cu-doped samples. An analysis of the changes in the *n(T)* dependencies for the irradiated Cu-doped samples upon isochronal annealing has shown that simultaneously with the enhanced disappearance of the A centres a strong decrease in the free carrier concentration measured at T = 300 K occurs. This decrease is associated with the introduction of deep acceptors having energy levels located lower than E_c – 0.45 eV in the gap. Maximum concentrations of these deep traps in the Cu-doped samples are much higher than the initial concentration of divacancy centres after irradiations and tend to approach the A centre concentration in the as-irradiated state. Fig. 2b shows that the disappearance of the divacancy occurs at lower annealing temperatures in the samples which were subjected to heat-treatments at 600 and 650 °C. An analysis of isochronal-annealing-induced changes in the conventional DLTS spectra as well as in the Laplace DLTS spectra has shown that the disappearance of the divacancy occurs through a reaction $V_2 + O_i \rightarrow V_2O$ in the irradiated as-grown Cz-Si sample, that is consistent with

the results of previous studies (11, 15). The formation of the V_2O complexes has not been observed in the irradiated Cu- and Ni-doped samples upon isochronal annealing (Fig. 3).

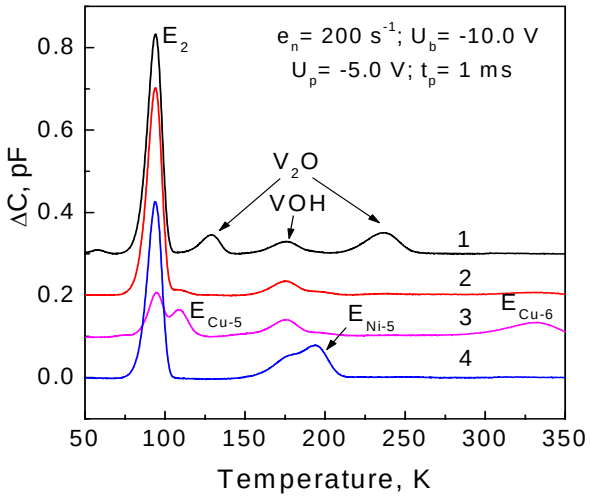


Figure 3. DLTS spectra of Cz-Si samples irradiated with 4 MeV electrons at 300 K and subjected to 30-min isochronal annealing with the last step at 275°C: (1) as-grown, (2) and (3) heat-treated at 600°C for 60 min without and with a Cu layer on the surface and (4) heat-treated at 650°C for 180 min with a Ni layer on the surface. The dose of irradiation was 1×10^{15} cm^{-2}. The spectra (1) to (3) are shifted on vertical axis for clarity.

It has been found from an analysis of the DLTS spectra for the irradiated Cu-doped samples that simultaneously with the disappearance of the VO and V_2 centres upon annealing an introduction of two other deep level traps, E_{Cu-5} and E_{Cu-6}, occurs (Fig. 3). The activation energies for electron emission from the E_{Cu-5} and E_{Cu-6} traps are 0.17 eV and 0.60 eV. An analysis of the changes in concentrations of the eliminated VO and V_2 centers and introduced E_{Cu-5} and E_{Cu-6} defects has allowed us to assign the later defects to acceptor levels of Cu-V_2 and Cu-VO complexes (16). It appears, that Cu atoms are released from some unidentified electrically inactive Cu containing complexes or precipitates upon anneals in the temperature range 100-200 ºC and interact effectively with vacancy-related defects, thus, causing significant decreases in the annealing out temperatures of the VO and V_2 and formation of stable Cu-V_2 and Cu-VO complexes.

Simultaneously with the disappearance of V_2 in irradiated Ni-doped samples a new trap, E_{Ni-5}, emerges (Fig. 3). The activation energy for electron emission for this trap is 0.365 eV. It can be suggested that the E_{Ni-5} trap is related to a complex of a radiation-induced defect and at least one Ni atom as the trap has been observed only in irradiated Ni-doped Si samples. However, we have not found clear anti-correlations in annealing-induced changes in concentrations of eliminated V_2 and A centres and introduced E_{Ni-5} traps. So, a microscopic structure of the E_{Ni-5} trap is not clear at the moment.

Results of Quantum-Chemical Calculations.

Possible structures of the Cu-VO and Ni-VO defect have been studied with the use of quantum chemical calculations within a frame of neglect of diatomic differential overlap (NDDO) approximation and PM5 parameterization (17). It is found that in the most stable configuration of the defect a Cu atom occupies a tetrahedral interstitial position nearest to the elongated Si-Si bond of the VO centre. The presence of the Cu atom results

Ex. 1016, Page 1045

in the further elongation of the Si-Si bond and a shift of the VO acceptor level to the middle of the gap. The results of calculations support the suggested assignment of the E_{Cu-6} trap to an acceptor level of Cu-VO complex. No stable structures incorporating a Ni atom and the VO centre have been found in our quantum chemical calculations.

Conclusions

The results obtained in the present work show that the enhanced disappearance of the VO and V_2 centres in electron-irradiated Cu-doped Cz-Si samples is associated with interactions of mobile Cu atoms with vacancy-related defects. Some of the resulting complexes are electrically active. Deep electron traps at $E_c - 0.17$ eV, and $E_c - 0.60$ eV have been assigned to acceptor levels of Cu-V_2 and Cu-VO complexes. According to results of quantum-chemical modelling, in the most stable configuration of the Cu-VO defect a Cu atom occupies a tetrahedral interstitial position nearest to the elongated Si-Si bond of the VO centre. In Si:Ni crystals an energy level at $E_c - 0.365$ eV is related to a complex incorporating at least one Ni atom and a radiation-induced defect.

References

1. A. A. Istratov and E. R. Weber, *Appl. Phys. A*, **66**, 123 (1998).
2. K. Graff, *Metal Impurities in Silicon-Device Fabrication*, Springer, Berlin (2000).
3. S. K. Estreicher, *Mater. Sci. in Semicond. Processing*, **7**, 101 (2004).
4. D. West, S. K. Estreicher, S. Knack, and J. Weber, *Phys. Rev. B*, **68**, 035210 (2003).
5. H. Lemke, *Phys. Stat. Solidi A*, **95**, 665 (1986).
6. S. D. Brotherton, J. R. Ayres, A. Gill, H. W. van Kesteren, and F. J. A. M. Greidanus, *J. Appl. Phys.*, **62**, 1827 (1987).
7. B. G. Svensson, M. O. Aboelfotoh, and J. L. Lindström, *Phys. Rev. Lett.*, **66**, 3028 (1991).
8. M. O. Aboelfotoh and B. G. Svensson, *Phys. Rev. B*, **52**, 2522 (1995).
9. W. Schröter and M. Seibt, in *Properties of Crystalline Silicon*, G. Hull, Editor, volume 20 of EMIS Data Review Series, chapter 9.11, INSPEC Publishing, London (1999).
10. L. Dobaczewski, A. R. Peaker, and K. Bonde Nielsen, *J. Appl. Phys.*, **96**, 4689 (2004).
11. V. P. Markevich, A. R. Peaker, S. B. Lastovskii, L. I. Murin, and J. L. Lindström, *J. Phys.: Condens. Matter*, **15**, S2779 (2003).
12. A. O. Evwarye and E. Sun, *J. Appl. Phys.*, **47**, 3776 (1976).
13. S. D. Brotherton and P. Bradley, *J. Appl. Phys.*, **53**, 5720 (1982).
14. J. L. Lindström, T. Hallberg, J. Hermansson, L. I. Murin, B. A. Komarov, V. P. Markevich, M. Kleverman, and B.G. Svensson, *Physica B*, **308-310**, 134 (2001).
15. E. V. Monakhov, B. S. Avset, A. Hallén, and B. G. Svensson, *Phys. Rev. B*, **65**, 233207 (2002).
16. V. P. Markevich, A. R. Peaker, I. F. Medvedeva, V. E. Gusakov, L. I. Murin, and B.G. Svensson, *Solid State Phenomena*, **131-133**, 363 (2008).
17. J. J. P. Stewart, *J. Comp. Chem.*, **10**, 209 (1989).

ECS Transactions, 18 (1) 1019-1022 (2009)
10.1149/1.3096566 © The Electrochemical Society

Effect of Oxygen Precipitates in Annealed Electron Irradiated Czochralski Silicon

Guifeng Chen[1], Xiaowei Ma[1], Yunna Bai[1], Lili Cai[1] and Yangxian Li[1]

[1] School of Material Science and Engineering, Hebei University of Technology, Tianjin, 300130, China
State Key Lab of Silicon Materials, Zhejiang University
Hangzhou 310027, People's Republic of China

The precipitation of oxygen has been investigated for 1.5 MeV electron irradiated silicon samples with the dose of 3.5×10^{17} e/cm^2. It has found that electron irradiation accelerated the oxygen precipitation in CZ-Si under higher annealing temperature. In addition, the effect of rapid thermal process (RTP) on the oxygen precipitation and on the denuded zone (DZ) formation in electron irradiated silicon wafers have been investigated by using optical microscope methods. The DZ can be formed at the surface in electron irradiated silicon wafers subjected to RTP pre-annealing. Furthermore, the width of DZ decreases with the RTP temperature.

Introduction

Oxygen precipitation in Czochralski (CZ) silicon used in microelectronic industry has been a subject of investigation for several decades. In the past days, much more research has been carried out in neutron irradiated Czochralski silicon but little in electron irradiated Czochralski silicon. Because the electron energy used for the irradiation (1.5MeV) is known to mainly cause the formation of point defects, the formation of clusters of vacancies and self-interstitials should be of no importance [1]. Due to the aggregation of these point defects, oxygen interstitials or combinations, extended defects form during the annealing and processing steps. In case that larger strain occurs around precipitates, punched-out dislocation loops (PDLs) will preferentially generate [2]. It is based on rapid thermal process (RTP) in which the oxygen precipitation behavior is controlled by the manipulation of vacancy rather than by oxygen concentration profiles [3].

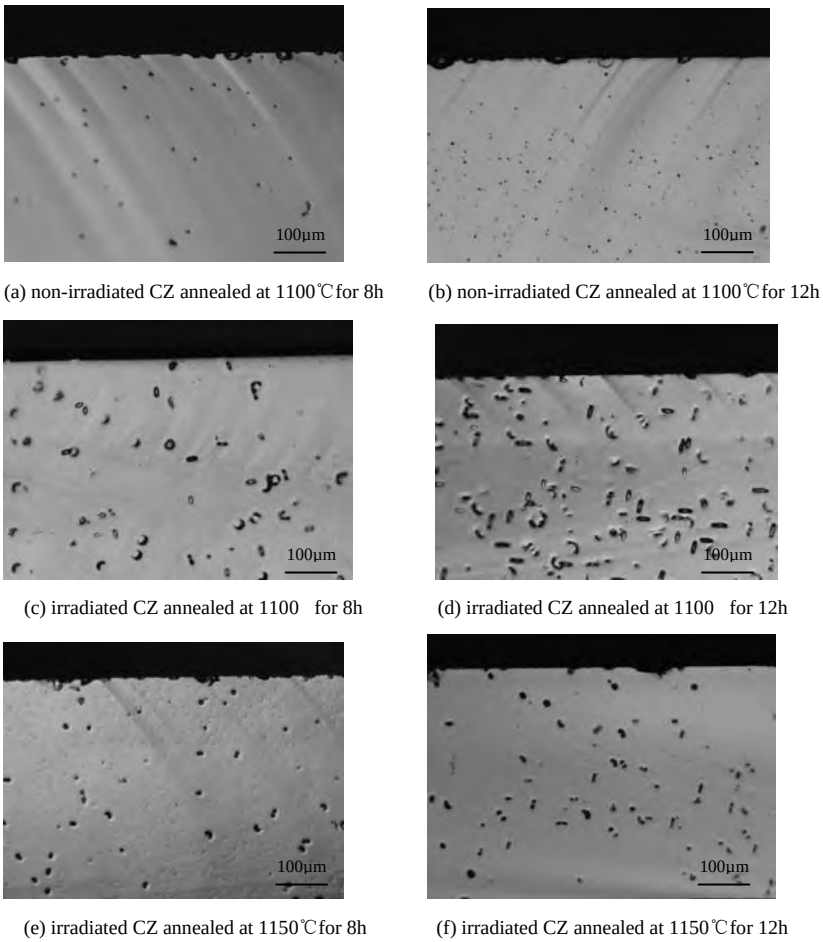
Experimental

The as-grown CZ-Si crystal is 4 in diameter, n type, <111> oriented and with an interstitial oxygen atoms concentration of 8×10^{17} atoms/cm^3. No dislocations in the as-grown CZ silicon were found. Parts of them were subjected to electron irradiation with a dose of 3.5×10^{17}e/cm^2. The samples with the thickness of about 600μm were chemically polished. Then all the samples were annealed at 1100 for 8 or 10h in argon atmosphere following RTP pre-annealing at different temperature with different cooling rate. The samples were preferentially etched in Wright etchant for 15 min. Finally, cross section of the wafers were inspected by means of an Olympus optical microscope.

Ex. 1016, Page 1047

Results and discussion

Fig. 1 shows the variation of BMD density with the annealing time at different higher temperature in irradiated and non-irradiated samples. A few of dot like etched pits in non-irradiated CZ silicon compared with large stacking faults and dislocation loops in irradiated samples could be detected. We attribute this effect to the form of defect cluster induced by electron irradiation, it may be as the core for the oxygen precipitation heterogeneous nuclei and those shorten the nucleation time and accelerate oxygen precipitates in the bulk of Si wafer [4]. However, in non-irradiated CZ silicon, the BMDs and oxygen precipitates are difficult to form because of lacking of heterogeneous nuclei [Fig. 1(a), (b)]. It can also be seen from Fig.1, the oxygen precipitation in all the investigated samples increases with annealing time. But it is worth noticing that the density of oxygen precipitates is lower at 1150 for 8h [Fig. 1(e), (f)]. It is considered that a number of oxygen precipitates nuclei related with irradiation have been dissociated. It has also found that an enhanced precipitation among irradiated samples appeared as a transient process [4].



(a) non-irradiated CZ annealed at 1100℃for 8h (b) non-irradiated CZ annealed at 1100℃for 12h

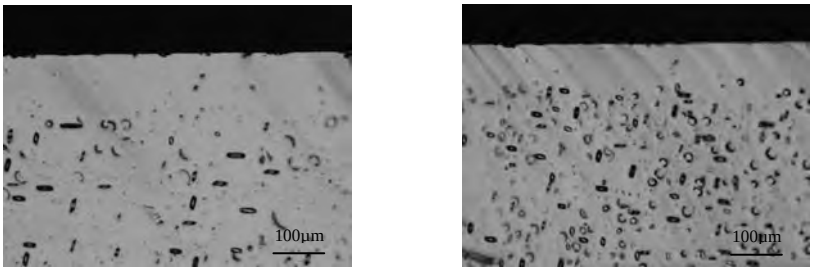
(c) irradiated CZ annealed at 1100 for 8h (d) irradiated CZ annealed at 1100 for 12h

(e) irradiated CZ annealed at 1150℃for 8h (f) irradiated CZ annealed at 1150℃for 12h

Fig. 1. The optical micrographs of BMDs in samples subjected to higher temperature as a function of time.

Fig.2 shows the density of BMDs in irradiated samples as a function of the annealing time at 1100 following RTP pre-annealing. Irradiated samples subjected to RTP annealing have larger number of micro-defects than that the samples without RTP pre-annealing [as Fig.1(c), (d)]. Furthermore, a perfect DZ can be formed in electron

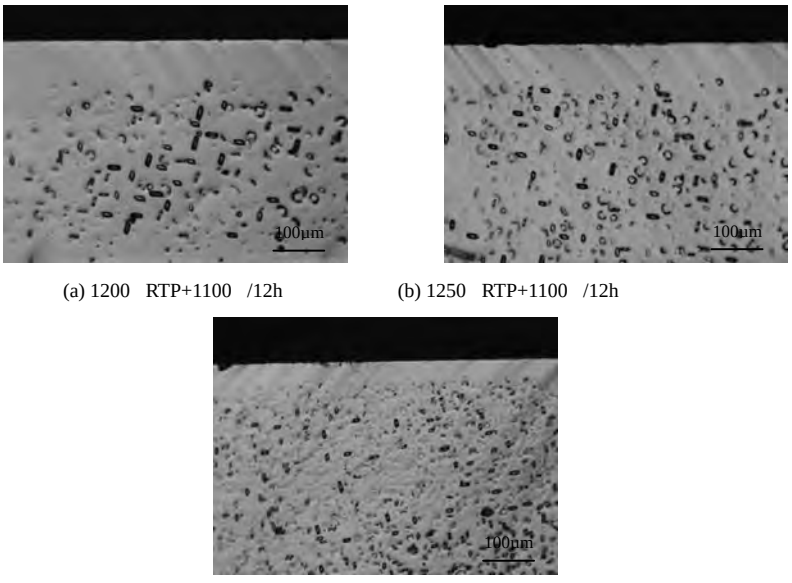
irradiated samples subjected RTP pre-annealing at near surface of the wafer. After RTP pre-annealing, a number of vacancies remain and form O_2V complexes with oxygen [5]. These vacancies greatly enhance the rate of oxygen clustering by providing free volume that reduces the energy necessary for the oxygen clusters to form. The nucleation of oxygen precipitates stops until the vacancies are consumed in the clustering process. The density of oxygen precipitates is determined by the vacancy concentration programmed into the bulk of the wafer and not by the concentration of oxygen [6]. As a result, a good DZ is formed at the surface and a large amount of BMD is present in the bulk of CZ wafer. The oxygen diffusion is enhanced by the presence of a large quantity of vacancies near the surface of the wafer [7].



(a) 1100 /8h (b) 1100 /12h

Fig. 2. Optical micrographs of the DZs and BMDs in the cross section of the irradiated CZ-Si wafers subjected to 1100℃, RTP was performed at 1200℃ for 50 s.

Fig.3 shows optical micrographs of cleavage plane in the irradiated samples as function of RTP temperature. From figure 3, it can also be observed that the DZ was



(a) 1200 RTP+1100 /12h (b) 1250 RTP+1100 /12h

（c）1280 RTP+1100 /12h

Fig.3. Optical micrographs of the DZs and BMDs in the cross section of the irradiated CZ-Si wafers subjected to annealing at 1100 for 12h following RTP with different temperature.

slightly narrower with the increase of RTP temperature, but the BMD density is higher. It is considered that vacancy concentration increases with the RTP temperature. Due to the competition mechanism of vacancies diffusion and generation, the generation of vacancies dominated over the out diffusion in the following annealing at higher temperature [8], which resulted in the shrinkage of the DZ in the irradiated samples

annealed at higher temperature. In addition, the size of stacking faults in the irradiated CZ–Si did not increase but decreased at higher RTP temperature. It is considered that dislocations could absorb self-interstitial Si atoms more easily [9]. Therefore once dislocations generate, they can absorb Si_I and lower the concentration of Si_I, so that stacking faults would shrink and dissolve.

Conclusion

The oxygen precipitation in the electron irradiation CZ-Si could be significantly accelerated by the action between oxygen and irradiation defects. Irradiation defects can act as the nuclei of oxygen precipitates and enhance oxygen precipitation. The DZs at the surface and denser BMD in the bulk can be formed in irradiated CZ samples subjected to RTP preannealing. The width of DZ decreases with the RTP temperature and the size of stacking faults in the irradiated CZ–Si decreased .It is considered that dislocation loops once formed can absorb the self-interstitial silicon atoms so that the generation of stacking faults is inhibited.

Acknowledgements

The part of this work is supported by the National Natural Science Foundation of China (50872028), Natural Science Foundation of Hebei Province (E2008000079) and Specialized Research Fund for the Doctoral Program of Higher Education (20050080006). We are also grateful for irradiating the samples by Tianjin Institute of Technical Physics.

References

1. Cao jianzhong, *Effects of irradiation in semiconductors*, Beijing, Science Press, (1993): 163-169.
2. Hongjie Wang, Deren Yang, Xuegong Yu, Xiangyang Ma, Daxi Tian, Yijun Shen, Liben Li, and Duanlin Que, Effect of oxygen precipitates and induced dislocations on oxidation induced stacking faults in nitrogen-doped Czochralski silicon, *Journal of Applied Physics*, 96[5]: 3031.
3. R. Falster, D. Gambaro, M. Olmo, M. Cornara, H. Korb, in: S. Ashok, J. Chevallier, K. Sumino, B.L. Sopori, Defect and Impurity Engineered Semiconductors and Devices II, Materials Research Society Symposium Proceedings, No. 510, *Materials Research Society*, Warrendale, PA, (1998), p. 27.
4. Tomas Hallberg, J. Lennart Lindstrom. Enhanced oxygen precipitation in electron irradiated silicon. *J.Appl. Phys.*, (1992), 72 (11):5130-5138.
5. M.A.katsuka, M.Okui, K.Sueoka, Effect of rapid thermal annealing on oxide precipitation behavior in silicon crystal, *Nuclear Instrument and Methods in Physics Research B*, (2002), 186: 46-54
6. Yu X, Yang D, Ma X, Que D. Effect of rapid thermal process on oxygen precipitation and denuded zone in nitrogen-doped silicon wafer, *Microelectronic Engineering*, (2003), 69:97-104.
7. T . Itoh, T. Abe, *Appl. Phys. Lett.*, 53 (1988) 523
8. M. Hourai, T. Nagashima, E. Kajita, S. Miki, T. Shigematsu, *J. Electrochem. Soc.*, 142 (1995) 3193.
9. J. Vanhellemont, *Appl. Phys. Lett*. 68 (1996) 3413.

ECS Transactions, 18 (1) 1023-1029 (2009)
10.1149/1.3096567 © The Electrochemical Society

Modeling Multi-Crystalline Silicon Growth in Directional Solidification Systems

Balaji Devulapalli[a] and Milind S. Kulkarni[a]

[a]Quantitative Silicon Research, MEMC Electronic Materials Inc.,
St. Peters, Missouri 63376, USA

The key to achieving high solar cell efficiency is to increase average grain size in silicon multicrystals and reduce dislocation density. The grain size and dislocation distributions in a silicon multicrystal are strongly influenced by the design of the solidification furnace and the operating conditions during solidification. This paper discusses a reported mathematical model that quantifies the temperature field in the furnace, silicon melt flow distribution, and the shape of the melt/solid interface, which influences the grain size and dislocations distributions. The global model described here takes into account all modes of heat transfer coupled with turbulent melt and inert gas flows, and can be used to manipulate the growth conditions to produce silicon multicrystals of higher qualities. The numerical approach discussed in this paper can be used to evaluate changes in equipment design and process conditions to increase the solidification rate with acceptable compromise in the multicrystal quality.

Introduction

Silicon in its pure multi-crystalline or monocrystalline form is the most widely used semiconductor material in solar cells and more than 80% of commercially installed PV modules are made from crystalline silicon (1). Directional solidification (DSS) is the leading commercial process for the multi-crystalline (mc) silicon growth because it offers lower manufacturing costs, higher throughput, and wider feedstock tolerance. The mc-silicon ingots are cast from the melt in a fused silica crucible inside a furnace. After complete melting of the feedstock, the solidification process starts at the crucible bottom and continues to the top by a controlled heat extraction through the melt/solid system to the bottom of the furnace. It is well known that the efficiency of solar cells depends on mc-silicon quality, which is largely determined by its temperature field during growth and the impurity distribution. Hence, the impurity control in silicon melt to reduce the grain size and the temperature field control are necessary to improve the yield of mc-silicon wafers.

Because of lack of instrumentation to directly observe inside the furnaces, and long process cycle times, it is convenient to apply numerical simulation to understand the temperature field, melt flow distribution and melt/crystal interface inside the furnace. Earlier numerical models did not consider convective heat transfer in the melt and simulated radiative heat transfer in a simplified manner (2). Liu et al. developed a transient global model for heat transfer, and performed calculations to study the temperature distribution, carbon segregation, and SiC particle precipitation in a cylindrical silicon ingot (3). The effects of the melt convection on temperature

distribution and evolution of the shape of the growth front have also been reported (4-6). Recently, Matsuo et al. studied the effects of crucible rotation on the distribution of oxygen concentration in a mc-silicon ingot (7). Wu et al. performed numerical simulations to compare different qualities of mc-silicon ingots grown in different furnaces, and reported that, the thermal stresses and the melt-crystal interface shape determine mc-silicon quality (8). Besides the temperature and melt flow distributions, the stress and dislocation generation were also investigated by both experiments and simulations (9, 10). Typically, silicon ingots with a weight of over 300 kg are produced and the Grashof number for these systems is large (10^8-10^9). Hence, the turbulence in the melt plays a critical role in determining the crystal quality. None of the published models so far have taken into account the effect of turbulent convection on the heat transfer in the melt. In this paper we present a comprehensive global heat transfer model that accounts for turbulent melt convection, inert gas flow and melt/crystal interface geometry. The global model provides a framework to investigate changes in the furnace design and process conditions to obtain a high quality crystal with reduced number of defects, at improved growth rates.

Model Description

Melt Flow Model

The global dynamics of the Directional Solidification of Silicon (DSS) process is characterized by solving the mass, momentum and energy conservation equations. The flow simulations are performed using the commercial CFD software FLUENT (11, 12). The code is pressure-based and uses an implicit RANS (Reynolds Averaged Navier Stokes) solver that employs a cell-centered finite volume scheme having second-order spatial accuracy. FLUENT discretizes the computational domain into elemental control volumes, and permits the use of quadrilateral/hexahedral, triangular/tetrahedral and hybrid meshes. The solver adopts “segregated” solution procedure, by which the transport equations are solved sequentially and the continuity equation is satisfied using the SIMPLE algorithm (12).

Melt flow in the DSS systems is complex because the buoyancy driven and surface tension driven flows exist simultaneously. Buoyancy forces resulting from the density variation with the temperature are taken into account by the Boussinesq approximation, and the surface tension effects are included by computing the Marangoni stresses at the melt/gas interface. The Grashof number for a typical system is in the range of 10^8-10^9, and the melt flow varies from transition to turbulent. A variation of k-ε turbulence model that accounts for the low-Reynolds number effects with appropriate wall functions is used to quantify turbulence (13). The heat transfer model incorporates the conductive heat transfer through all solids, coupled with the convective heat transfer in the melt and gas, and the radiation between all solids and liquid surfaces in the DSS furnace. The radiative heat transfer is calculated using a discrete ordinate method described in detail in Ref (14).

Solidification Model

The solidification process in Fluent is modeled using the enthalpy-porosity formulation. The essential feature of this formulation for the region undergoing phase change is that

the latent heat effects are isolated in a source term in the energy equation. In a system undergoing a phase change, the total enthalpy is expressed as

$$H = h + L_{mix} \quad [1]$$

$$h = h_{ref} + \int_{T_{ref}}^{T} C_p dT \quad [2]$$

For pure systems, the latent heat contribution, L_{mix}, can be written in terms of the latent heat of material, L

$$\left.\begin{array}{ll} L_{mix} = \beta L & \text{for T} > \text{T}_{melt} \\ = 0 & \text{for T} < \text{T}_{melt} \end{array}\right\} \quad [3]$$

where β is the liquid fraction and can vary between 0 (for a solid) and 1 (for a liquid). With the definition of the enthalpy provided by equation [1] the energy equation can be written as

$$\frac{\partial}{\partial t}(\rho h) + div(\rho v h) = div(K \nabla T) + S_h \quad [4]$$

where h is the sensible enthalpy, ρ is the density, v is the melt velocity and S_h is the latent heat source term.

To address numerical challenges, the liquid fraction is effectively updated according to the following scheme, with the constraint that it is between 0 and 1 (both values inclusive):

$$\beta_{i+1} = \beta_i + kC_p(T_i - T_{melt})/L \quad [5]$$

where k includes relaxation factor and other numerical terms, and the subscript i denotes the ith iteration. Equation [5] closely mimics the conditions set by Equation [3], without causing difficult numerical problems. The enthalpy-porosity formulation treats the partially solidified region as a porous medium, and assumes that the flow in this region is governed by Darcy's law. The porosity in each numerical cell is set equal to the liquid fraction in that cell. As the porosity decreases, the velocities in the partially solidified region also decrease, down to a limiting value of zero, when it becomes completely solid. The momentum sink resulting by the reduction in the porosity in this region takes the following form:

$$S = C\frac{(1-\beta)^2}{(\beta^3 + \varepsilon)}v \quad [5]$$

where C is a constant depend on the morphology of porous media, and ε is a small number to prevent division by zero. C is taken as a large number for the so-called isothermal solidification, representing the monocomponent systems. Sink terms are also added to k-ε equations to account for the presence of the solidified region. Additional details regarding the numerical implementation are reported by Voller et al. (15).

Results and Discussion

The geometry of a DSS furnace along with the temperature distribution is shown in Figure 1. The furnace operates at a low pressure in general, using argon as the purge gas. During the growth phase, the side insulation is moved up to allow heat extraction from

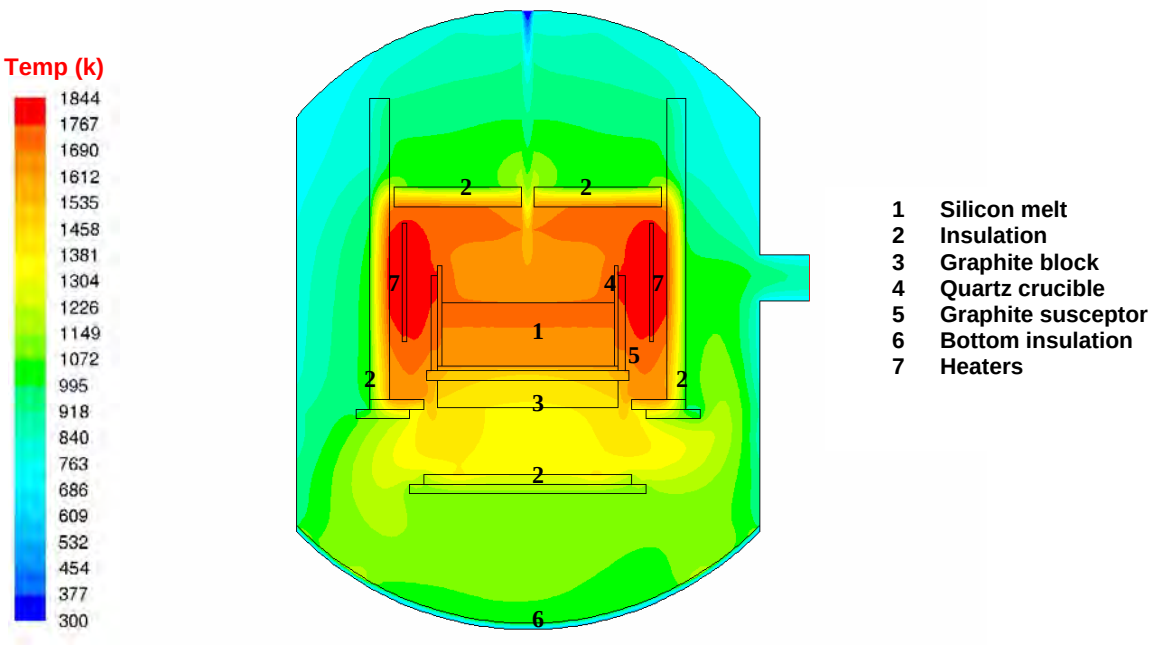


Figure 1. Global temperature distributions inside a DSS furnace

the bottom while the crucible remains stationary. The critical aspect of the global heat transfer calculations of the DSS system is the prediction of the melt/solid interface shapes. Figure 2 (a-d) shows the predicted melt/crystal interface shapes and the streamlines during different stages of solidification.

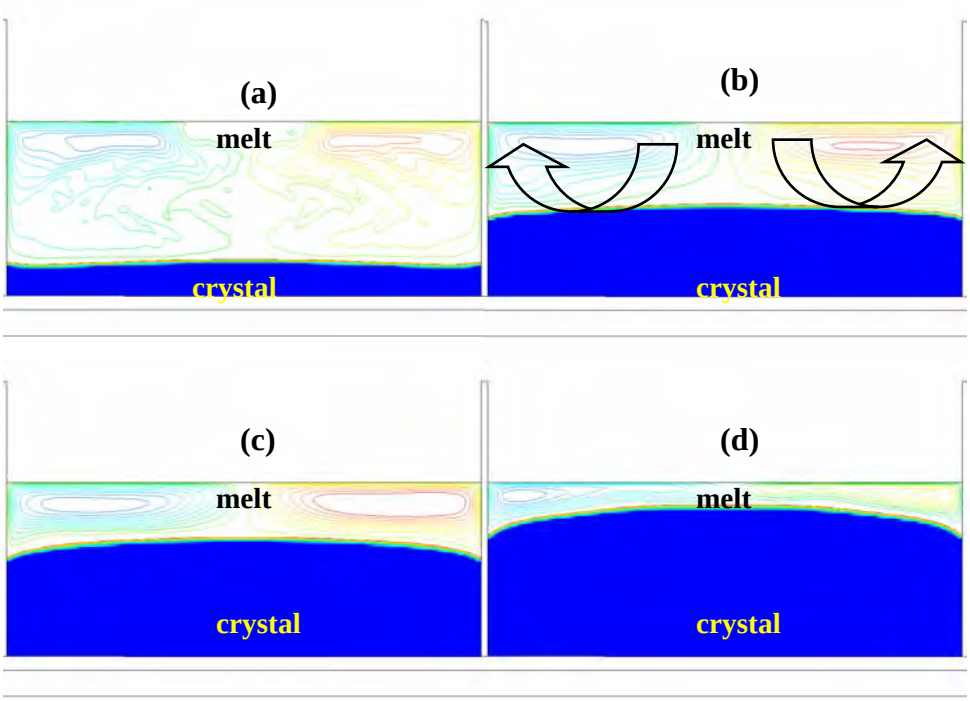


Figure 2. Predicted melt-crystal interface shapes and distributions of the streamlines in silicon melt at different stages of solidification (a) t = 2 hrs (b) t = 6 hrs (c) t = 10 hrs (d) t = 14 hrs

The bulk of the melt flow is primarily driven by the buoyancy. The results show two large vortices or cells that slide along the interface and flow upward along the crucible walls. The flow near the free surface is also aided by the weak Marangoni stresses. The simulations indicate that the buoyancy effects are weaker during initial stages of solidification, when the temperature gradients in the melt are small, causing instabilities in the melt flow marked by the irregular and time-dependent streamlines, as shown in Figure 2 (a). The flow instabilities vanish in the melt, however, as the temperature gradients and buoyancy forces increase at later stages of solidification, as shown in Figure 2 (b-d). Understanding the driving forces of these flows provides insights into how heat is transferred within the melt and how the melt/crystal interface shape can be controlled in the DSS systems.

The predicted temperature distributions in the melt and solid at different stages of solidification are shown in Figure 3 (a–d). The isotherms change their shape along height

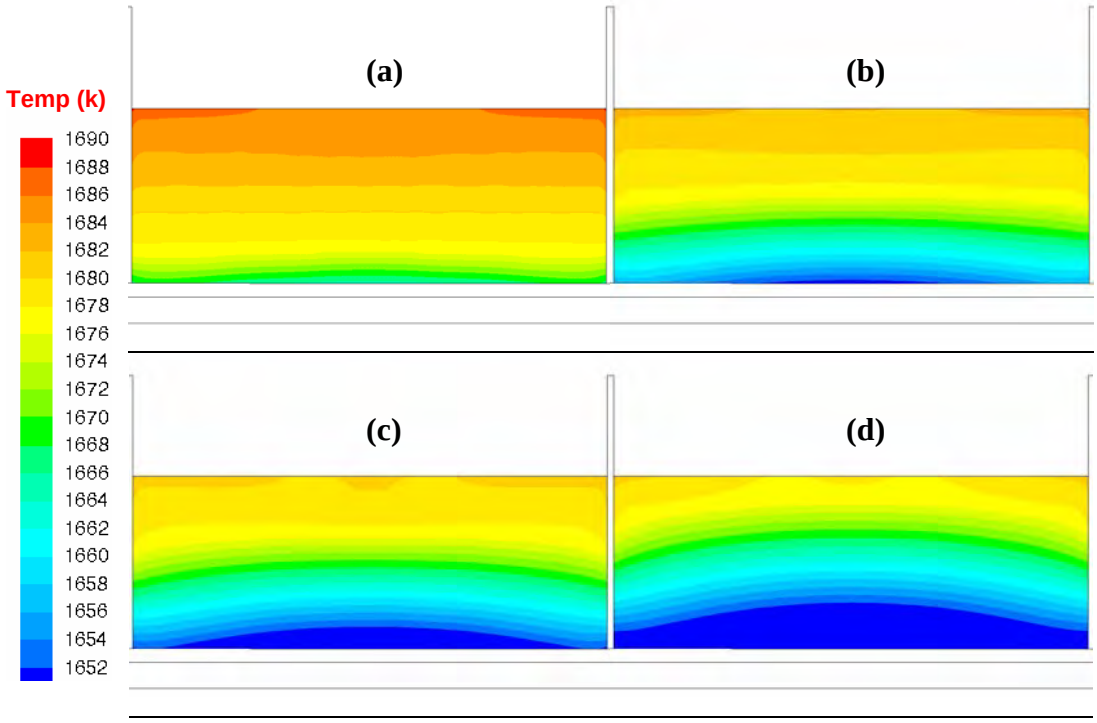


Figure 3. Predicted temperature distributions of silicon melt and solid at different stages of solidification (a) t = 2 hrs (b) t = 6 hrs (c) t = 10 hrs (d) t = 14 hrs

of the solid and melt. During early stages of solidification, the melt/solid interface and the isotherms near the interface are approximately flat. This is because of the balanced effect of the side heaters and the colder environment at the bottom. However, as solidification progresses, the melt temperatures throughout silicon are higher near the crucible walls than in the middle because of the increased effect of the heater at the upper portion of the crucible. The simulations indicate that at the beginning of solidification, the isotherms are practically flat in shape and change to the concave shape to the melt side during the later phases of the growth, as shown in Figures 3 (a-d). The melt and crystal temperature fields, which are coupled with the melt flow field, determine the melt/solid interface shape, and the solidification proceeds normal to the interface.

Moreover, the predicted temperature fields in the crystal can be used to calculate the stress fields, which directly affect the distribution of dislocations.

For mc-silicon based solar cells, the size distribution of grains and the dislocation density within the grains affect the solar cell efficiency. It is known that the stresses induced by the temperature gradients and the distribution of impurities such as carbon, oxygen, nitrogen and residual metals in the ingot are important factors contributing to the formation of defects. Control of the melt temperature field, the melt flow field and the curvature of the melt/crystal interface are critical in controlling the impurity distributions both in the melt and in the solidified ingot. A relatively flat melt/crystal interface is necessary to reduce the thermal stresses and to improve the uniformity of the radial distribution of impurities in the silicon melt. Sustaining a flat melt/crystal interface throughout the solidification process leads to larger grains and lower dislocations densities, thus improving the efficiency of solar cells.

Conclusions

Multi-crystalline (mc) silicon ingots grown by the directional solidification process contain several defects such as grain boundaries and dislocations that degrade the solar cell efficiency, and the major challenge is to achieve a high directional solidification rate with reduced number of defects. Computational models provide an effective tool to understand the physics underlying the ingot growth. A self-consistent global heat transfer model for the directional solidification systems, which takes into account all modes of heat transfer coupled with the turbulent melt convection, and the evolution of the melt/crystal interface, is solved. The results provides insights into the cause and effect relationships among melt flow field, temperature field, interface shape and impurity distribution in steady-state as well as transient systems. The model described here provides a computational framework to investigate changes in the furnace design/process parameters and identify key growth conditions that affect the formation of grain boundaries and dislocations in the growing ingots. Furthermore, the evolution of the temperature field can be used as an input to compute thermal stresses during solidification, which directly affect the ingot quality.

References

1. A. Luque and S. Hegedus, Editors, p. 153, Handbook of Photovoltaic Science and Engineering, Wiley, New York (2003).
2. D. Franke, T. Rettelbach, C. Habler, W. Koch and A. Muller, *Solar Energy Materials & Solar Cells*, **72**, 83 (2002).
3. L. Liu, S. Nakano and K. Kakimoto, J. *Crystal Growth,* **310**, 2192 (2008).
4. A.T. Kuliev, N. V. Durnev and V. V. Kalaev, *J. Crystal Growth*, **303**, 236 (2007).
5. E. A. Meese, E. J. Ovrelid, H. Laux and M. M. Hamdi, 19th European Photovoltaic Solar Energy conference, p. 793 (2004).
6. D. Vizman, J. Friedrich and G. Muller, *J. Crystal Growth*, **303**, 231 (2007).
7. H. Matsuo, R. B. Ganesh, S. Nakano, L. Liu, K. Arafune, Y. Ohshita, M. Yamaguchi and K. Kakimoto, The 5th International Symposium on Advanced Science and Technology of Silicon Materials, p. 135 (2008)
8. B. Wu, N. Stoddard, R. Ma and R. Clark, *J. Crystal Growth*, **310**, 2178 (2008).

9. M. M. Hamdi, E. A. Meese, H. Laux, and E. J. Ovrelid, *Materials Science Forum*, **67**, 447 (2006).
10. X. Chen, S. Nakano, L. Liu and K. Kakimoto, The 5th International Symposium on Advanced Science and Technology of Silicon Materials, p. 124 (2008)
11. S. R. Mathur, and J.Y. Murthy, *Numerical Heat Transfer*, **31,** 195 (1997).
12. S. E. Kim, S. R. Mathur, J.Y. Murthy, D. Choudhury, *AIAA* (1997).
13. D. Choudhury, Introduction to Renormalization Group Method and Turbulence Modeling, Fluent Inc. Technical Memorandum (1993).
14. J. Y. Murthy and S. R. Mathur, *AIAA 98-0860* (1998).
15. V. R. Voller, M. Cross and N. C. Markatos, *Intl. J. for Numerical Methods in Engineering*, **24**, 271 (1987).

ECS Transactions, 18 (1) 1031-1035 (2009)
10.1149/1.3096568 © The Electrochemical Society

Simulation Analysis of Point Defects in a Silicon Ingot during a Unidirectional Solidification Process for Solar Cells

X. J. Chen[a], S. Nakano[a], L. J. Liu[b] and K. Kakimoto[a]

[a] Research Institute for Applied Mechanics, Kyushu University, 6-1 Kasuga-koen, Kasuga, Fukuoka 816-8580, Japan
[b] National Engineering Research Center for Fluid Machinery and Compressors, School of Energy and Power Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

A transient global model was used to obtain the solution of a thermal field within the entire furnace of a unidirectional solidification process. The melt-solid interface shape was obtained by a dynamic interface tracking method. Then, based on the global solution of heat transfer, the effects of growth rate V_g, temperature gradient G and ratio V_g/G on point defects were analyzed. Finally, several different melt-solid interface shapes were obtained by using different solidification times. Then the effects of solidification time on ratio V_g/G and point defects were also studied.

Introduction

Worldwide solar photovoltaic market installations have been rapidly increasing in recent years, and the unidirectional solidification method is an important cost-effective technique for large-scale production of multi-crystalline silicon material. It has been reported that quasi-single crystals can also be grown by the unidirectional solidification method by BP solar (1). Micro-defects formed by self interstitials and vacancies in a silicon ingot during growth from the melt of crystals can greatly degrade the quality of a silicon ingot for highly efficient solar cell materials. The growth rate V_g and temperature gradient G affect point defects in silicon ingot, and the type and concentration of point defects are determined by the ratio V_g/G (2). It is therefore necessary to simulate the solidification process with a transient global model and to elucidate the effects of growth rate V_g, temperature gradient G and ratio V_g/G in a silicon ingot on point defects during the solidification process.

Simulation model and method

To study the ratio V_g/G in a silicon ingot during the unidirectional solidification process, an axisymmetric furnace shown schematically in Figure 1 is considered. Figure 1 shows the configuration, domain partition and computational grid of a unidirectional solidification furnace. The growth system consists of silicon melt, silicon ingot, crucible, pedestal, insulation shield and heater. The domains of all components in the furnace are subdivided into 15 block regions, and each block region is covered by a structured grid.

For calculation of heat transfer and melt flow, the following major assumptions are made: (a) the geometry of the furnace configuration is axisymmetric, (b) radiative transfer can be modeled as diffuse-gray surface radiation, (c) the effect of gas flow in the

furnace is negligible, and (d) the melt flow in the crucible is laminar and incompressible. Thus, the governing equations for flow, conductive heat transfer and radiative heat transfer under the above assumptions can be expressed as follows:

$$\nabla \cdot \vec{V} = 0, \quad [1]$$

$$\rho\left(\frac{\partial \vec{V}}{\partial t} + \vec{V} \cdot \nabla \vec{V}\right) = -\nabla p + \nabla \cdot \left[\mu\left(\nabla \vec{V} + \nabla \vec{V}^{T}\right)\right] - \rho \vec{g} \beta_T \left(T - T_m\right), \quad [2]$$

$$\rho c\left(\frac{\partial T}{\partial t} + \vec{V} \cdot \nabla T\right) = \nabla \cdot \left(k \nabla T\right), \quad [3]$$

$$\frac{q(\vec{x})}{\varepsilon(\vec{x})} \int_{\vec{x}' \in \partial V} K(\vec{x}, \vec{x}') \frac{1-\varepsilon(\vec{x}')}{\varepsilon(\vec{x}')} q(\vec{x}') dS' = \sigma T^4(\vec{x}) - \int_{\vec{x} \in \partial V} K(\vec{x}, \vec{x}') \sigma T^4(\vec{x}') dS', \quad [4]$$

where $\vec{V}$, ρ, p, μ, $\vec{g}$, β_T, c and k represent melt velocity, melt density, melt pressure, melt viscosity, gravitational acceleration, thermal expansion coefficient, heat capacity and thermal conductivity, respectively, and $q(\vec{x})$, $\varepsilon(\vec{x})$, $K(\vec{x}, \vec{x}')$ and σ are heat flux, radiative emissivity, view factor and Stefan-Boltzmann constant, respectively. The symbols $\vec{x}$ and $\vec{x}'$ are infinitesimal radiative surface elements on the enclosure, dS is the area of the infinitesimal surface element $\vec{x}$, and $K(\vec{x}, \vec{x}')$ is the surface view factor between $\vec{x}$ and $\vec{x}'$, which is described in detail in Ref. (3).

The conductive heat transfer in all solid components, radiative heat exchange between all diffuse surfaces in the whole furnace, and Navier-Stokes equations for the melt flow are coupled and solved by a transient global model, and the melt-solid interface shape is obtained by a dynamic interface tracking method (3-5).

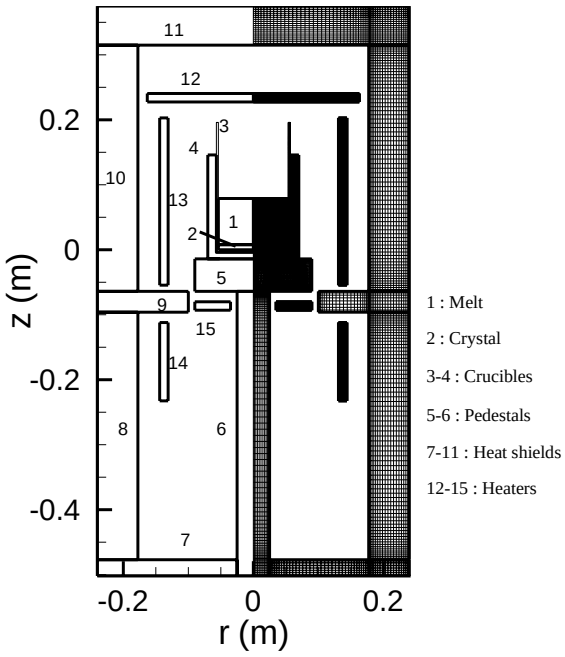


Figure 1. Configuration, domain partition (left part) and computation grid (right part) of a unidirectional solidification furnace.

Results and discussion

It has been reported that the type and concentration of point defects that remain in a crystal depend on the ratio V_g/G (vacancies if $V_g/G > \xi_i$, interstitials if $V_g/G < \xi_i$, where ξ_i is a certain constant) (2). The time histories of growth rate, temperature gradient in the growth direction and ratio V_g/G at the center line of the silicon crystal for three different solidification processes are shown in Figures 2 to 4. The solidification times are about 15, 10 and 5 hours for Figures 2, 3 and 4, respectively. As shown in the figures, there are some oscillations in growth rate at the beginning of each of the three solidification processes because of the instability of melt flow. For the case in Figure 2, since the solid fraction is less than 0.20, the Rayleigh number of melt flow is larger than 1.0×10^5 (6) and the convective flows in the melt become unsteady (time-dependent). After that, as the temperature difference in melt and height of melt are reduced, the melt flows become steady and then the growth rate becomes smooth.

As the figures show, the temperature gradient G decreases as the melt is solidified, and growth rate V_g and ratio V_g/G increases. It was pointed out in reference (2) that the critical value ξ_i is about $3.3 \times 10^{-5}\, cm^2/sK$, and the type of point defects in the crystal during the solidification process could therefore be determined by using the figures. For a longer solidification time, 15 hours, the ratio V_g/G is less than $3.3 \times 10^{-5}\, cm^2/sK$ during the entire solidification process, so there are only interstitials in the center of the crystal during the solidification process. However, as the solidification time is decreased, the ratio V_g/G increases at each solid fraction because of increased growth rate. Figures 3 and 4 indicate that the point defects would be vacancies after the solid fraction are about 0.7 and 0.3 for solidification times of 10 and 5 hours, respectively, because the ratio V_g/G is larger than $3.3 \times 10^{-5}\, cm^2/sK$ after that solid fraction.

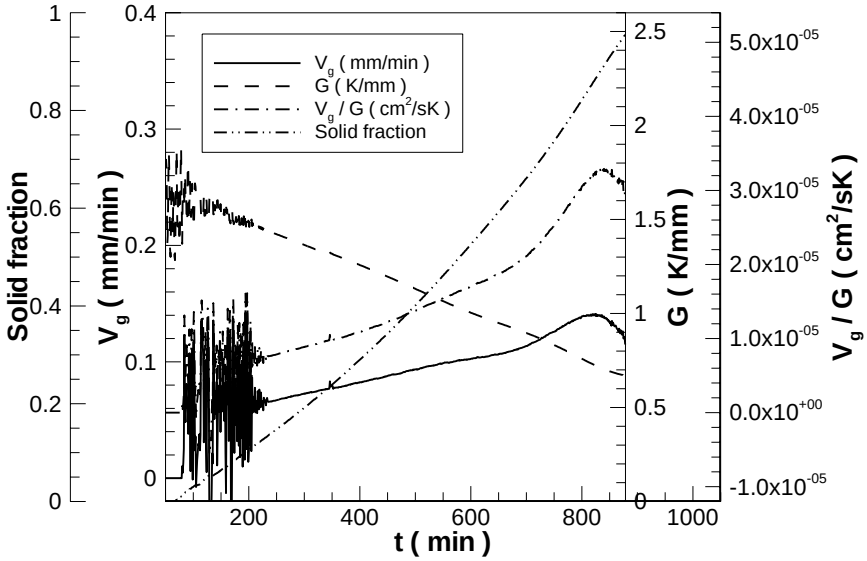


Figure 2. Growth rate, temperature gradient and ratio V_g/G as a function of time during a solidification process (solidification time of 15 hours).

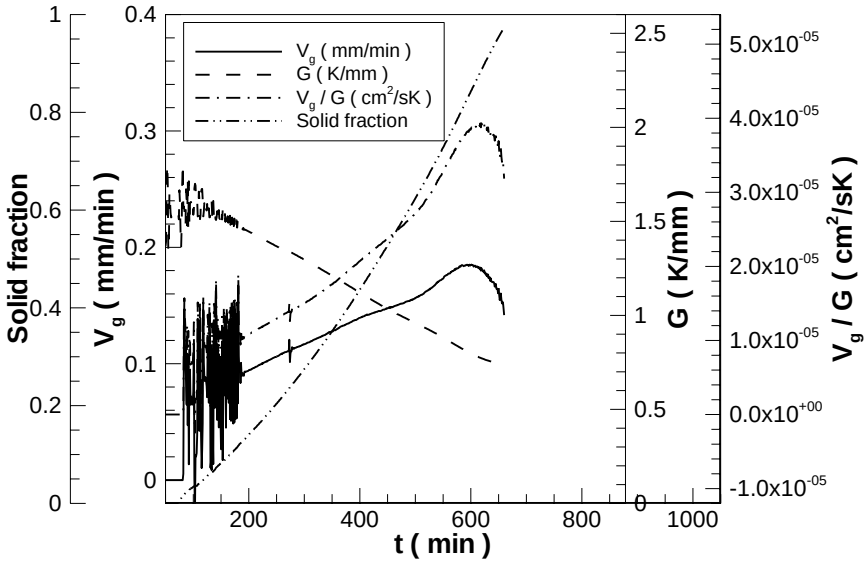


Figure 3. Growth rate, temperature gradient and ratio V_g/G as a function of time during a solidification process (solidification time of 10 hours).

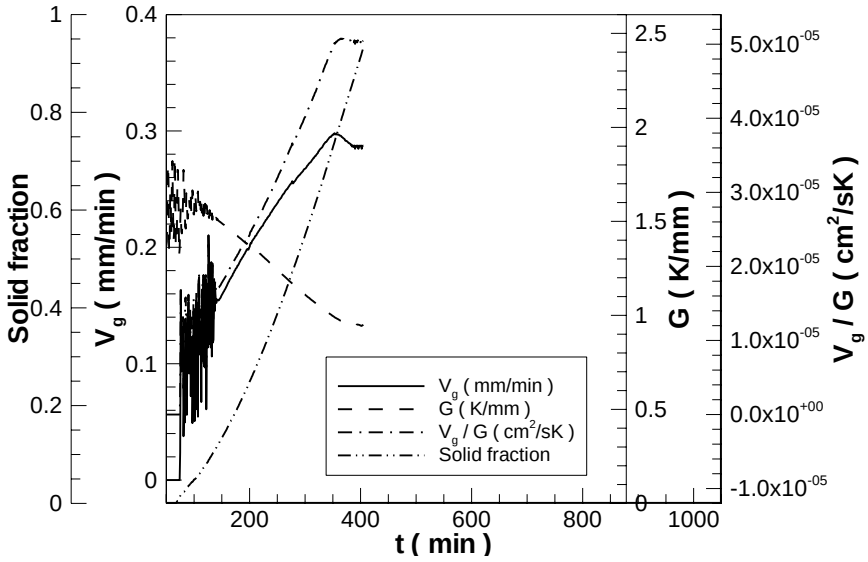


Figure 4. Growth rate, temperature gradient and ratio V_g/G as a function of time during a solidification process (solidification time of 5 hours).

Figure 5 shows the ratio V_g/G along the solid-melt interface after solidification for 5, 10 and 15 hours. As the figure shows, the ratio V_g/G increases as the solidification time is decreased, the reason for this being that the growth rate is increased as the solidification time is reduced. The point defects near the solid-melt interface would be vacancies when the solidification time is less than 10 hours, because the ratio V_g/G along the interface after the solidification process is larger than $3.3 \times 10^{-5}\ cm^2/sK$ for shorter solidification time. The figure also shows that the ratio V_g/G near the crucible wall is larger than that in the center of crystal. For a longer solidification time, 15 hours, point defects near the crucible wall would be vacancies and those in the center part would be interstitials. The radius of the boundary of vacancies and interstitials is about 45 *mm*.

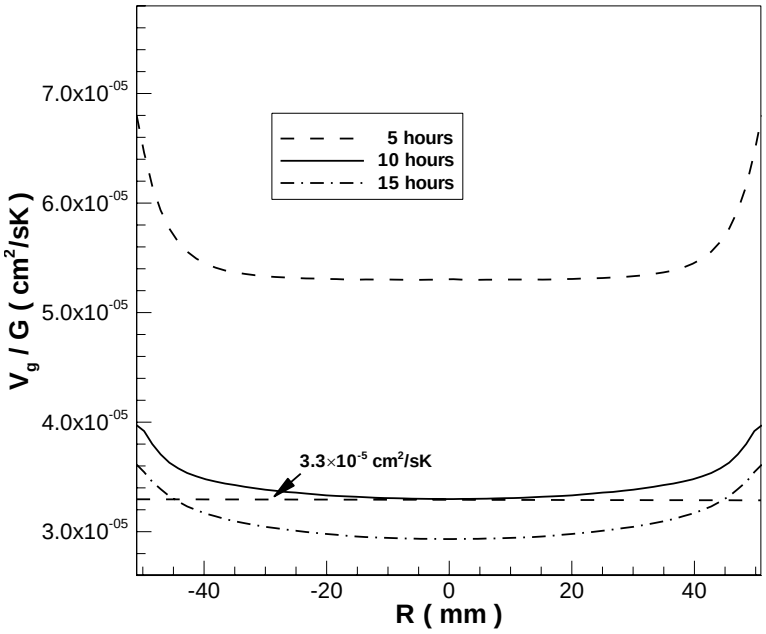


Figure 5. Ratio V_g/G along the solid-melt interface after solidification for different times.

Conclusion

Based on the global solution of heat transfer, the effects of growth rate V_g, temperature gradient G and ratio V_g/G on point defects were analyzed. The ratio V_g/G was increased as the melt solidified. The point defects at the center of the crystal would be interstitials during solidification for 15 hours and would be vacancies after solid fractions become 0.7 and 0.3 for solidification times of 10 and 5 hours, respectively. The V_g/G ratios for three different solidification times indicated that the point defects at the top of crystal would be vacancies for a shorter solidification time.

References

1. N. Stoddard, B. Wu, L. Maisano, R. Russell, J. Creager, R. Clark and J. M. Fernandez, Proceedings of 18th Workshop on Crystalline Silicon Solar Cells & Modules: Materials and Processes p. 7, Colorado (2008).
2. V. Voronkov, *J. Crystal Growth*, **59**, 625 (1982).
3. L. Liu, K. Kakimoto, *Int. J. Heat Mass Transfer*, **48**, 4481 (2005).
4. L. Liu, S. Nakano and K. Kakimoto, *J. Crystal Growth*, **292**, 515 (2006).
5. L. Liu, S. Nakano and K. Kakimoto, *J. Crystal Growth*, **310**, 2192 (2008).
6. D. Hurle (Ed.), *Handbook of Crystal Growth*, vol. 2, Elsevier, Amsterdam (1994).

ECS Transactions, 18 (1) 1037-1042 (2009)
10.1149/1.3096569 © The Electrochemical Society

Distributions of Light Elements and Their Precipitations Grown by Unidirectional Solidification Method in Multicrystalline Silicon for Solar Cells

H. Matsuo[a], S. Hisamatsu[a], Y. Kangawa[a, b], and K. Kakimoto[a, b]

[a] Graduate School of Engineering, Kyushu University, 816-8580, Japan
[b] Research Institute for Applied Mechanics, Kyushu University, 816-8580, Japan

We investigated the relationship between the formation of small grains and distribution of precipitates such as Si_3N_4 and SiON in a multicrystalline silicon ingot grown by the unidirectional solidification method. Si_3N_4 precipitate was mainly observed inside the small grain region, while SiON was mainly precipitated outside there. We found that SiON started to precipitate prior to formation of Si_3N_4 precipitates. Therefore, precipitation of SiON prior to precipitation of Si_3N_4 plays a key role in the formation of small grains.

Introduction

The present multicrystalline silicon substrates have a significantly larger grain size (>1 cm), which is sufficiently large compared with the minority carrier diffusion length (1). Therefore, the grain boundaries in the typical cases have a negligible effect on conversion efficiency (1). However, it is reported that small grains of less than 1 mm in size are accidentally formed in a multicrystalline silicon ingot (2). Small grain region in a multicrystalline silicon ingot decreases conversion efficiency of solar cells (3). Arafune et al. reported that small grain region included large amount of nitrogen (2). It has been reported that nitrogen in the melt precipitated as Si_3N_4 when concentrations of impurities exceeded its solubility limits (4). Reimann et al. reported that nitrogen concentration in the small grain region is lower than those outside the small grain region (5). This means that Si_3N_4 is precipitated in the small grain region. It has also been reported that Si_3N_4 reacts with oxygen in the melt and forms silicon oxynitride (SiON) (6). Therefore, oxygen concentration also affects the formation of precipitates.

From this point of view, we studied the relationship between formation of small grains and precipitations in a silicon ingot by using Fourier transform infrared spectrometry (FTIR).

Experimental method

Boron-doped multicrystalline silicon has been grown by the unidirectional solidification method (7). The ingot was cut parallel to the growth direction into slices of 0.5 mm in thickness. Each sample was etched and polished before measurement by FTIR to remove the damaged layer formed by the cutting process. The size of the infrared beam was set to 50 μm with a square shape. The resolution and number of accumulation of the spectrum were set to 4 cm^{-1} and 182 times, respectively.

The absorbance of interstitial oxygen was measured at 1106 cm^{-1} (8). The silicon vibration peak was subtracted by using a floating zone silicon wafer. The concentration of oxygen has been determined using the calibration factor of 3.14 ×

TABLE I. Categorized SiON compounds.

Compound	Wavenumber (cm^{-1})
$SiO_{0.19}N_{1.21}$	840 - 880
$SiO_{0.36}N_{1.1}$	880 - 950
$SiO_{0.62}N_{0.92}$	950 - 1025
$SiO_{1.4}N_{0.4}$	1025 - 1080

Ex. 1016, Page 1065

10^{17} cm^{-2} (9). The concentration of Si_3N_4 was estimated from the main peak at 840 cm^{-1} (10). SiON peak position depends on the composition between oxygen and nitrogen. In this paper, we categorized peak position as shown in TABLE I. This categorization is based on a report by Teschner et al. (11). We showed related concentrations of Si_3N_4 and SiON because there are no calibration coefficients of them.

Distributions of impurities and precipitation in the whole ingot

First, we studied the kind of precipitates formed in the ingot. Figure 1 shows typical FTIR spectrum obtained from the ingot. One can see there are three peaks around 840, 925-940 and 1106 cm^{-1} for Si_3N_4, $SiO_{0.36}N_{1.1}$ and interstitial oxygen, respectively. Related to the Si-O-N precipitates, only the $SiO_{0.36}N_{1.1}$, i.e., peaks around 925-940 cm^{-1}, was observed (see TABLE I.).

Next, measured distributions of impurity and precipitate concentrations in the whole ingot and compared them with distributions of small grains. Figures 2 (a), (b), (c) and (d) show a photograph of the vertically sliced multicrystalline silicon and distributions of oxygen, Si_3N_4 and $SiO_{0.36}N_{1.1}$ concentrations, respectively. The left and right sides of the picture correspond to the center and crucible wall, respectively, as shown in Fig. 2 (a). The area of measurement of impurity and precipitate concentrations by using FTIR is indicated by white dashed lines. The measurement interval was set to 5 mm. Small grains of less than 1 mm in size were observed near the center of the ingot in Fig. 2 (a). Black dashed lines in the figures indicate

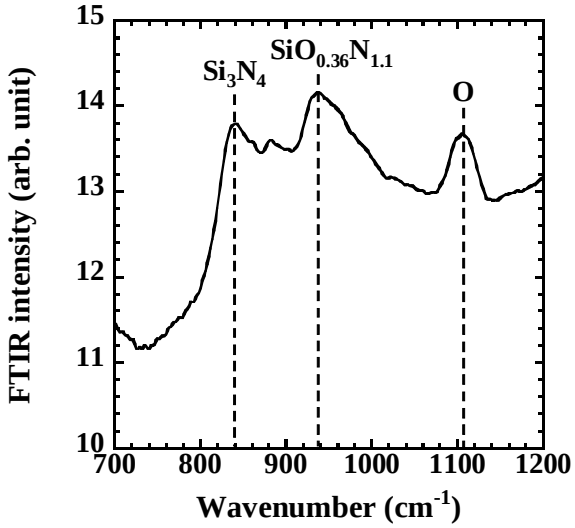


Figure 1. FTIR spectrum in a multicrystalline silicon ingot.

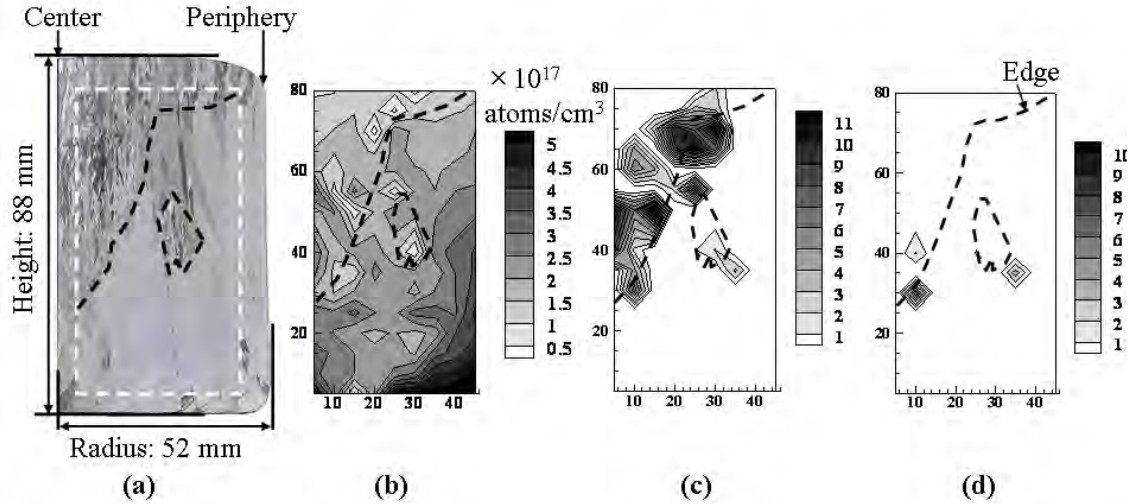


Figure 2. Distributions of impurity and precipitate concentrations obtained by FTIR measurements in the whole ingot. Photograph of the vertically sliced multicrystalline silicon wafer (a) and distributions of oxygen (b), Si_3N_4 (c), $SiO_{0.36}N_{1.1}$ concentrations (d), respectively.

the boundary between with and without the small grain. In Fig. 2(d), we can see $SiO_{0.36}N_{1.1}$ was precipitated near the edge of small grain region. Precipitations of Si_3N_4 were found in and/or near the small grain region as shown in Fig. 1 (c). Moreover, Si_3N_4 concentration was increased with increase in height from the bottom of the ingot. However, precipitation of Si_3N_4 was not observed at the center near the top of the ingot while small grains exist in this region. These results revealed that precipitate of $SiO_{0.36}N_{1.1}$ was only observed near edge of small grain region, while Si_3N_4 precipitated in the small grain region in the multicrystalline silicon ingot. The results suggest that formation of precipitates near the edge of small grain region plays a key role of formation of small grains. In order to clarify the formation mechanisms of small grain region, it seems important to investigate the detailed distributions of impurity and precipitate concentrations near the edge of the small grain region at the middle part of the ingot in Fig. 2 (a) with a spatial resolution of 2 mm.

Distributions of impurities and precipitation near the edge of small grain region

In order to discuss the formation mechanisms of small grain region, we measured detailed distributions of impurity and precipitate concentrations near the edge of small grain region. The measurement interval was set to 2 mm. The positions of the measurements were from 32 to 44 mm in the growth direction from the bottom of the ingot and from 1 to 41 mm in the radial direction from center of the ingot.

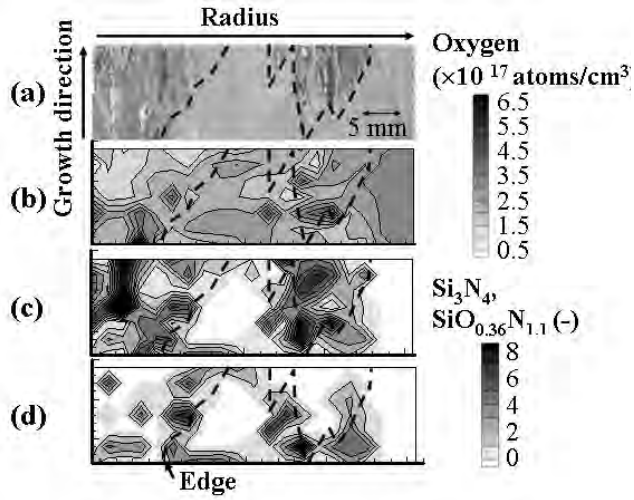


Figure 3. Distributions of impurity and precipitate concentrations obtained by measurements in 2 mm resolution at the middle of the ingot. Photograph of the vertically sliced multicrystalline silicon wafer (a) and distributions of oxygen (b), Si_3N_4 (c), $SiO_{0.36}N_{1.1}$ concentrations (d), respectively.

Figures 3 (a), (b), (c) and (d) show a photograph of the vertically sliced multicrystalline silicon and distributions of oxygen, Si_3N_4 and $SiO_{0.36}N_{1.1}$ concentrations, respectively. A dashed line indicates the edge of the small grain region in these figures. Oxygen concentrations in the center of the ingot at the top of Fig. 3 (b) were lower than compared with that at bottom of the figure. However, we could not find a clear relationship between impurity concentrations and formation of small grains. Si_3N_4 precipitate was mainly observed inside the small grain region, while $SiO_{0.36}N_{1.1}$ was mainly precipitated outside there. The results imply that $SiO_{0.36}N_{1.1}$ precipitated prior to Si_3N_4.

Precipitates process during solidification

From above discussion, we estimated that $SiO_{0.36}N_{1.1}$ started to precipitate prior to precipitation of Si_3N_4. In order to confirm the feasibility of this speculation, we discuss precipitates process by using a phase diagram of the Si-N-O system in this section.

Figure 4 shows a phase diagram of the Si-N-O system at 1685 K in a silicon crystal (6). In this figure, Si_2N_2O precipitate region is shown. The composition of Si_2N_2O is similar to the measured composition, i.e., $SiO_{0.36}N_{1.1}$as shown in Fig. 2(d) and Fig. 3(d). This results support our experimental results.

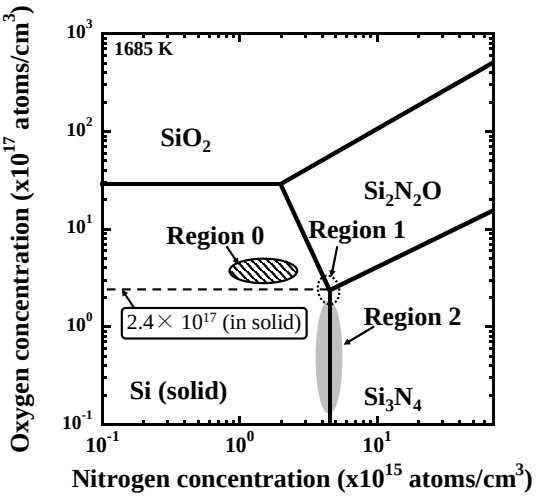


Figure 4. Phase diagram of the Si-N-O system in a solid at 1685 K. (6)

As you can see in Fig. 4, type of precipitates are related to the chemical composition, i.e., oxygen and nitrogen concentrations in the melt. Therefore, we discuss the change in oxygen and nitrogen composition during growth.Oxygen and nitrogen concentrations were in Region 0 shown in Fig. 4 at the initial stage of growth. Oxygen concentration was decreased from the bottom to the top of the ingot as shown in Fig. 2(b) because oxygen in the melt evaporates from the top of the melt to gas phase as SiO gas (12). Nitrogen concentration was increased with increase in crystal height from the bottom to the top of the ingot since the segregation coefficient of nitrogen was reported to be 0.0007 (13). Therefore, chemical composition during the solidification is thought to be changed from Region 0 to Region 2 via Region 1 in Fig. 4. Consequently, $SiO_{0.36}N_{1.1}$ precipitated prior to Si_3N_4. In order to confirm these assumptions, we focused on oxygen concentration in the region where $SiO_{0.36}N_{1.1}$ and Si_3N_4 precipitated.

When $SiO_{0.36}N_{1.1}$ was precipitated, interstitial oxygen was consumed because oxygen reacts with nitrogen and silicon. In order to confirm the relationship between oxygen consumption and $SiO_{0.36}N_{1.1}$ precipitation, we investigated oxygen concentration in the region where $SiO_{0.36}N_{1.1}$ precipitated. Figures 5 (a) and (b) show the relationship between

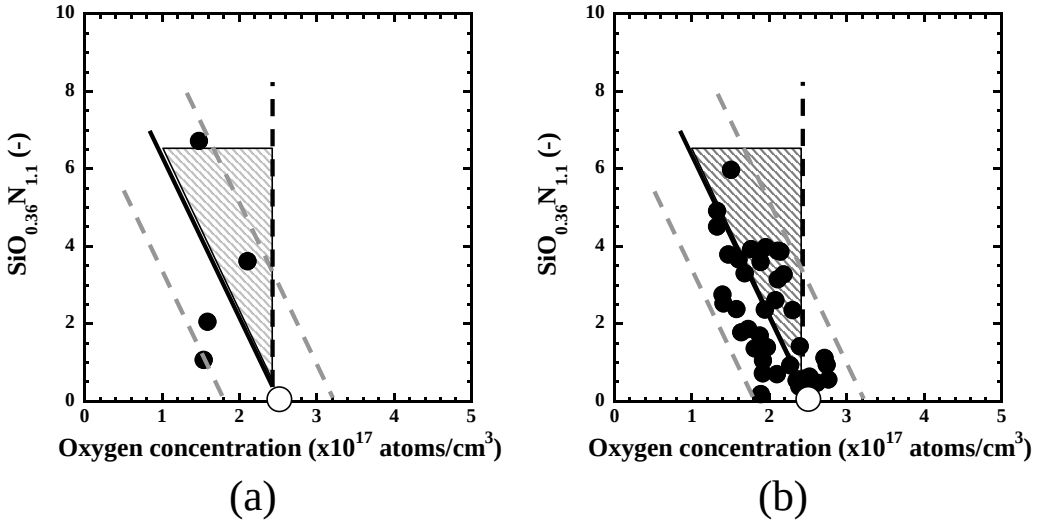


Figure 5. Relationship between oxygen and $SiO_{0.36}N_{1.1}$ concentration in a multicrystalline silicon wafer. Data of whole ingot measurements (a) and data of 2 mm spatial resolution (b). Abscissa and ordinate show oxygen and $SiO_{0.36}N_{1.1}$ concentrations, respectively. The shaded area corresponds to amount of consumed oxygen.

oxygen and $SiO_{0.36}N_{1.1}$ concentrations in multicrystalline silicon. Figures 5 (a) and (b) were plotted by using the data of whole ingot measurements shown in Figs. 2 (b) and (d) and the data shown in Figs. 3 (b) and (d), respectively. It can be seen that oxygen concentration linearly decreased with increase in $SiO_{0.36}N_{1.1}$ concentration. We estimated the initial oxygen concentration before precipitation of $SiO_{0.36}N_{1.1}$ as shown by a white circle in Fig. 5. The initial oxygen concentration in Fig. 5 is approximately 2.5×10^{17} atoms/cm^3. The initial oxygen concentration is almost the same as that in Region 1 shown in Fig. 4. While, an oxygen atom reacts with two atoms of nitrogen and silicon and they form $SiO_{0.36}N_{1.1}$. A precipitated $SiO_{0.36}N_{1.1}$ atom corresponds to a consumed oxygen atom. Therefore, the gradient indicated by black solid line in Fig. 5 showed the relationship between oxygen consumption and $SiO_{0.36}N_{1.1}$ precipitation. These results suggest that the following process occurred during crystal growth. Oxygen and nitrogen concentrations become saturated to the values shown in Region 1 in Fig. 4. Consequently, $SiO_{0.36}N_{1.1}$ was precipitated, and oxygen was consumed as shown in Fig. 5. The chemical composition when $SiO_{0.36}N_{1.1}$ was precipitated corresponds to Region 1 in Fig. 4.

Precipitation of Si_3N_4 was observed after precipitation of $SiO_{0.36}N_{1.1}$ at the top of the ingot as shown in Figs. 2 (c) and (d). We estimated that oxygen concentration in the region where Si_3N_4 precipitated was less than 2.4×10^{17} atoms/cm^3 in the ingot from the phase diagram of Si-N-O shown in Fig. 4. Figures 6 (a) and (b) show the relationship between oxygen and Si_3N_4 concentrations in the multicrystalline silicon. Figures 6 (a) and (b) were plotted by using the data of whole ingot measurements shown in Figs. 2 (b) and (c) and the data shown in Figs. 3 (b) and (c), respectively. Oxygen concentration has no relationship with Si_3N_4 concentration. Moreover, oxygen concentration in the region where Si_3N_4 precipitated was less than 2.4×10^{17} atoms/cm^3 in the ingot. Such chemical composition corresponds to Region 2 indicated by gray ellipse in Fig. 4. It can be seen that some data exceed the value of 2.4×10^{17} atoms/cm^3 as shown in Fig. 6 (b). This is because of fitting error of data processing during FTIR measurement.

From these results, formation of $SiO_{0.36}N_{1.1}$ started to precipitate prior to precipitation of Si_3N_4. We found that small grains start to form near the interface of precipitates between Si_3N_4 and $SiO_{0.36}N_{1.1}$. Therefore, precipitation of $SiO_{0.36}N_{1.1}$ prior to precipitation of Si_3N_4 plays a key role in the formation of small grains.

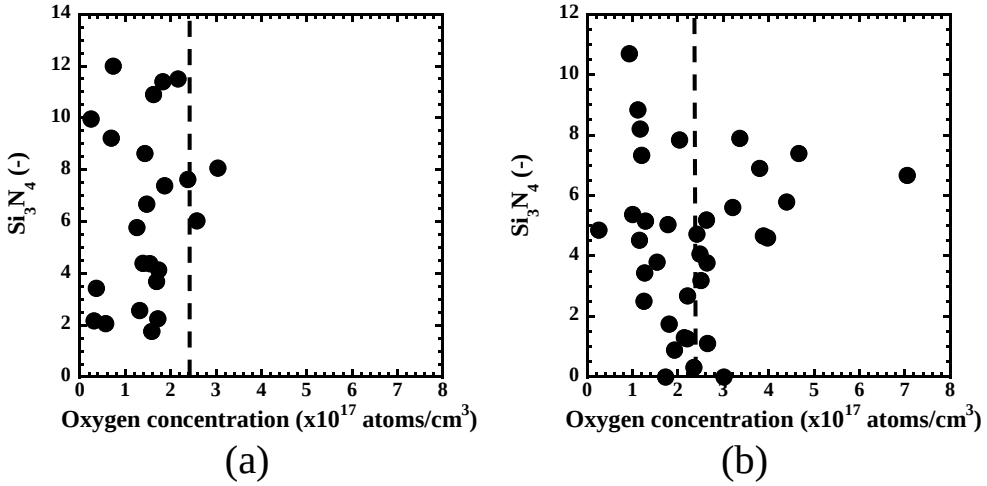


Figure 6. Relationship between oxygen and Si_3N_4 concentrations in a multicrystalline silicon wafer. Data of whole ingot measurements (a) and data of 2 mm spatial resolution (b). Abscissa and ordinate show oxygen and Si_3N_4 concentrations, respectively.

Summary

We investigated the relationship between formation of small grains and distribution of precipitates such as Si_3N_4 and SiON in a multicrystalline silicon ingot grown by the unidirectional solidification method. Si_3N_4 precipitate was mainly observed inside the small grain region, while SiON was mainly precipitated outside there. Oxygen concentration decreased with increase in SiON concentration. This is because oxygen was consumed to precipitate SiON. Moreover, oxygen concentration has no relationship with Si_3N_4 concentration. Oxygen concentration decreases from the bottom to the top of the ingot. Nitrogen concentration was increased with increase in crystal height from the bottom to the top of the ingot due to segregation phenomena. Consequently, SiON started to precipitate prior to the formation of Si_3N_4 precipitate. These precipitation processes can also be discussed using a Si-O-N phase diagram. We found that small grains formed near the interface of precipitations between Si_3N_4 and SiON. Therefore, precipitation of SiON prior to precipitation of Si_3N_4 plays a key role in the formation of small grains.

Acknowledgments

The authors would like to thank professor Arafune in University of Hyogo, professor Yamaguchi and Ohshita in Toyota Technological Institute for preparing the ingot.

References

1. K. Arafune, T. Sasaki, F. Wakabayashi, Y. Terada, Y. Ohshita and M. Yamaguchi, Physica, B 376 (2006) 236.
2. K. Arafune,Y. Ohshita and M. Yamaguchi, *Proceedings of THE FORUM ON THE SCIENCE AND TECHNOLOGY OF SILICON MATERIALS 2007*, 346 (2007).
3. M. Imaizumi, T. Ito, M. Yamaguchi and K. Kaneko, *J. Appl. Phys.*, **81**, 7635 (1997).
4. H. Laux, Y. Ladam, K. Tang, E.A. Meese, *Proceedings of 21st EUPVSEC*, 4 (2006).
5. C. Reimann, T. Jung, J. Friedrich and G. Müller, *Proceedings of the 33rd IEEE PVSC*, (2008).
6. J. Lu, G. Rozgonyi, J. Rand and R. Jonczyk, *J. Crystal Growth*, **269**, 599 (2004).
7. K. Arafune, E. Ohishi, H. Sai, Y. Ohshita and M. Yamaguchi, *J. Crystal Growth*, **308**, 5 (2007).
8. H. J. Möller, C. Funke, A. Lawerenz, S. Riedle and M. Werner, *Sol. Energy Mater. Sol. Cells*, **72**, 403 (2002).
9. T. Iizuka, S. Takasu, M. Tajima, T. Arai, T. Nozaki, N. Inoue and M. Watanabe, *J. Electrochem. Soc.*, **132**, 1707 (1985).
10. E. Theodossiu, H. Baumann, W. Matz, A. Mücklich, *Phys. Stat. Sol.*, **194**, 47 (2002).
11. U. Teschner, *Phys. Stat. Sol. (a)*, **121**, 641 (1990).
12. K. Hoshikawa and X. Huang, *Mater. Sci. Eng.*, B **72**, 73 (2000).
13. Y. Yatsurugi, N. Akiyama, Y. Endo and T. Nozaki, *J. Electrochem. Soc.*, **120**, 975 (1973).

ECS Transactions, 18 (1) 1043-1047 (2009)
10.1149/1.3096570 © The Electrochemical Society

Numerical Simulation of Thermal Distribution during the Directional Solidification of Polycrystalline Silicon Ingot

Dan Wu, Weizhong Sun, Qiuyan Hao, Jianguo Zhao and Caichi Liu

Department of material, HeBei University of Technology, Tianjin, Hongqiao 300130, CHINA

Abstract: Directional solidification is one of the key technologies of multi-crystalline silicon ingots in commercial photovoltaic market. During the directional solidification process, the changes of technologic parameters will influence the temperature gradient and the shape of the solid–liquid interface, which determine the qualities of polycrystalline ingots. In this paper, the relationship between the crucible lower rates with the temperature gradient and the shape of the solid–liquid interface were studied by numerical simulations. The results showed that when the crucible lower rate is less than 60mm/h , the temperature gradient is getting larger with the increase of crucible lower rate, and the shape of solid–liquid interface remains convex. But when the crucible lower rate is higher than 60mm/h, the shape of solid-liquid interface changes into concave.

1 Introduction

In photovoltaic (PV) market, about two-third of the solar cells were made of multi-crystalline silicon (mc-Si) substrates because the production cost of mc-Si is lower than that of single crystalline silicon (sc-Si)(1,2). Directional solidification of multi-crystalline silicon is the industrial production technique with the largest market share for solar silicon of about 60%. During the ingot casting process, the driving force of growth is coming from the supercooling of the melt(ΔT) .In order to make sure the high quality of polycrystalline silicon ingot, a steady and perfect thermal distribution in the casting process are prerequisite. A perfect thermal distribution makes the Si melt not only retain the melting state by assimilating the heat from the external heater but also release the redundance thermal from the plane of growing in the way of conduction and radiation(3). However, in order to make the directional solidification continue, it is a key to control the thermal conductor among them, the shape of solid-liquid interface is a very important aspect too, because it affects the forming of material structure and the phase directly, which determine material performance finally (4). If we want to control the thermal distribution of directional solidification, crucible lower rate and cooling rate are two important respects. In this paper, base on the theory of directional solidification, the influence of different crucible low rates, cooling rates and cooling materials on the thermal distribution during the directional solidification of multi-crystalline silicon ingot casting process was simulated by using the finite element method.

Ex. 1016, Page 1071

2. The simulation of thermal distribution during the directional solidification

2.1 Establishment of model

In this paper, the Bridgeman method to produce multi-crystalline silicon was simulated. In order to achieve the directional solidification process, the temperature gradient at the solid-liquid boundary is higher than zero by controlling the rate of crucible lower and the flow of cooling source. The sketch of casting furnace was shown in Fig.1. The used ceramic crucible is 10mm in wall thickness and 690×690×420mm^3 in size. The heater is made of graphite, and Ar is as the shielding gas.

Ihe whole simulation process was supposed as this, (1) Solidification is expected to begin when the polysilicon melted completely; (2) Conduction heat transfer only occurs in the solid state and melt state;(3) the polysilicon density does not change with temperature during the solidification process; (4) the material properties related to the calculation are isotropic;(5) the process of solidification is under a relatively balanced condition without considering the flow of solid-liquid phase;(6) the flow of molten polysilicon in the crucible was considered as an incompressible laminar flow.

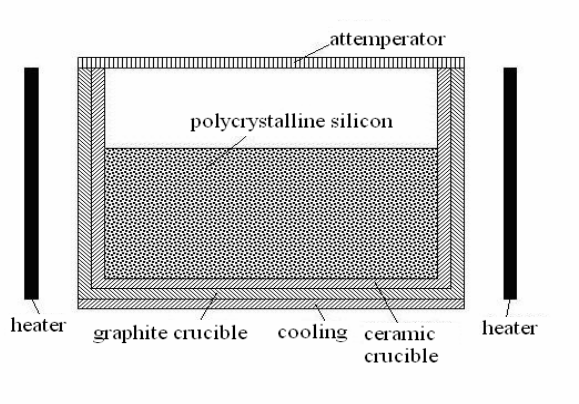


Figure 1. Sketch of the casting furnace

2.2 Simulation parameters

The thermal distribution of polysilicon directional solidification was simulated by finite element method. The representational parameters used in the simulation were listed in table1, and material characteristics were showed in table2.

TABLE I. Physical characteristics of Argon.

Physical Characteristics	**Value**	**.**
Conductivity /W·m^{-1}·K^{-1}	0.2373	x
Density/kg·m^{-3}	0.558	x
Heat capacity /J·kg^{-1}·K^{-1}	520.8	x
Viscosity /kg·m^{-1}·s^{-1}	4.95×10^{-6}	

TABLE . Material Characteristics

Material Character	**Cp/j· kg^{-1}· k^{-1}**	**Conductivity/W· m^{-1}· K^{-1}**
polycrystalline Silicon ingot	900	24
polycrystalline Silicon molten	1000	67
Ceramic crucible	3900	3.2
Graphic crucible	2100	41.9
felt cover	836	0.15

3.1 Simulation Results

The different crucible lower rates (48 mm/h, 51 mm/h, 54 mm/h, 57 mm/h and 60 mm/h) were used to analyze the thermal distribution of multi-crystalline silicon ingots. The temperature of the heater, the melt and the cooling water was 1800K, 1690K and 300K respectively in this process. All of those parameters were used to calculate 1200 times to get the thermal field distribution in different crucible lower rate, which was shown in Figure.2.

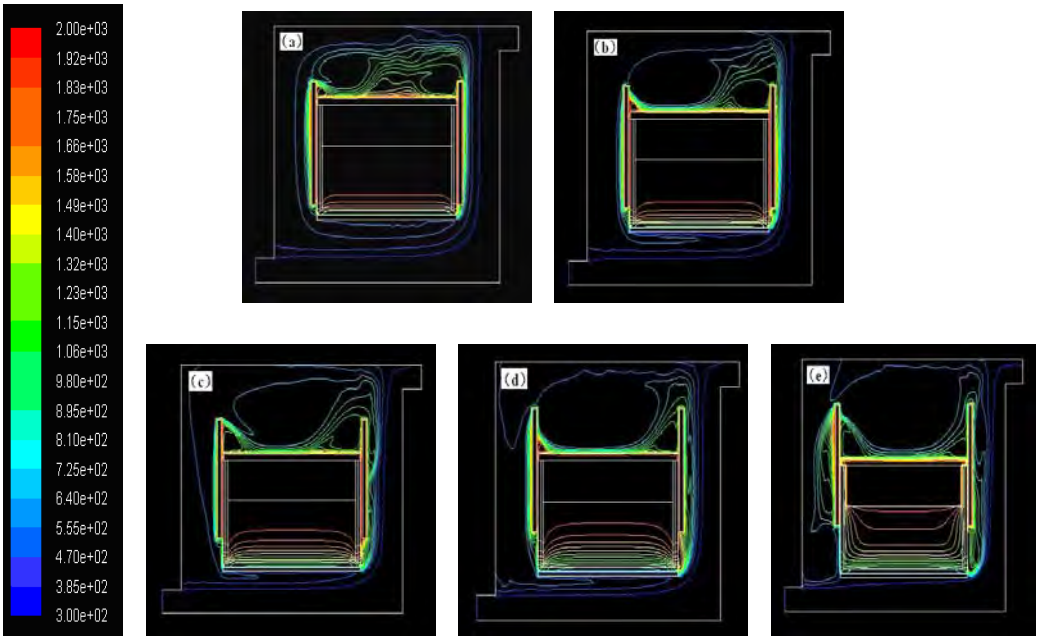


Figure.2 The influence of the different crucible lower rate on the solid-liquid interface

Figure.2 (a), (b), (c),(d) showed the relationship between crucible lower rate and the shape of the solid-liquid boundary. From the fig.2 it can be seen that only when the crucible lower rate is less than 60mm/h, the shape of solid-liquid boundary is convex, otherwise the shape is concave. Different crucible lower rates lead the different temperature gradients and the moving up of the solid-liquid boundary in the crucible, then make the polysilicon melt solidify constantly. The growing rate of the multi-crystalline silicon ingots is controlled by the temperature gradient of the crucible, which is made by melt (high-temperature region) and crystals (low-temperature region). The density of temperature isothermals can show the size of the temperature gradients. The sparser temperature isothermals indicates the smaller temperature gradient, the denser temperature contour indicates the larger temperature gradient. Thus it can be found that

the temperature gradient increased with crucible lower rate by the temperature contour in Figure.2.

Figure.3 shows the temperature change curve of the different height in the crucible, which were measured in the condition of different crucible lower rate. The temperature curve is the most smooth and is the minimum affected when the crucible lower rate is 48㎜/h. The temperature curve is steeped with the increase of the crucible lower rate, and when the crucible lower rate is 64mm/h, this effect will reach the maximum value. It can also be found in Fig.3 that with the increase of the crucible lower rate the temperature that both in melt and in crystal will gradually decrease, the temperature gradient will gradually increase, and the temperature contour and the temperature gradient in crystal will change more significantly than that in melt.

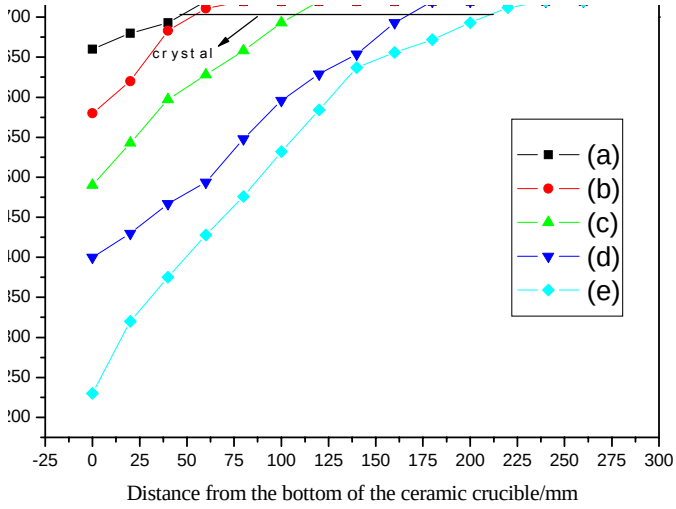


Figure.3 the influence of crucible lower rate on the temperature distribution in the ceramic crucible

4. Analysis and discussion

At the process of multi-crystalline silicon fabrication, the reasonable distribution of thermal field not only offers the driving force of crystal growth but also makes the crystal growth in high quality manner. The control of the solid-liquid interface shape and the temperature gradient distribution are very important to the distribution of thermal field. The influence of crucible lower rate on the solid-liquid interface and temperature gradient was discuss as follows.

4.1 The influence of crucible lower rate on the solid-liquid interface

In the process of crystal growth, the crystal with flatter growth interface has better quality, but in the actual production the shape of concave or convexity frequently was used to control the interface shape. At the process of multi-silicon ingot, the convexity interface must be used to keep crystal grow(5). The reason is that the concave interface is easy to form fine-grain region, where the grain boundary and dislocation growing in. Both of grain boundary and dislocation, which will be the electron-hole recombination centre, consequently, reduce the conversion efficiency of solar cell. Thus, keeping the solid-liquid interface of melt convex will be beneficial to get the columnar crystal, so as to get purity, tissue dense and zero defect ingots.

It can be found in Figure.2 that the solid-liquid interface of melt keeps convex when the crucible lower rate is less than 60mm/h and when the crucible lower rate is higher than 60mm/h the interface becomes concave.

4.2 The influence of crucible lower rate, cooling rate and cooling material on the temperature gradient

With the decrease of crucible lower rate, polysilicon melt is removed away from the heat source, the temperature of the crystal decreased, but the temperature gradient of crystal and melt increase, so the crystal growth rate will accelerate. The faster crystal growth is better to inhibit composition under cooling and to improve the quality of the crystal[6]. Thus, keep increasing the crucible lower rate is benefic to get high-quality ingots.

5. Conclusions

At the process of multi-silicon by Bridgeman method, the distribution of thermal field during the directional solidification were simulated and analyzed by using different crucible lower rates, the conclusions are obtained as follows.

(1) Crucible lower rate is the most important factor to the shape of solid-liquid interface, when the crucible lower rate is less than 60mm/h, the interface keeps convexity, which is beneficial to get high quality silicon ingots.

(2) The larger crucible lower rate was used, higher temperature gradient in crystal and melt was obtained, and the crystal grows more quickly.

Acknowledgments

The subject was supported by the National Nature and Science Foundation of China (50572022).

References

1. D.Sati and R.Einhaus, Energy Materials and Solar Cells, 72(2002) 27.
2. R.Kvande, O.Mjos and B.Ryningen, Materials Science and Engineering A, 413-414(2005) 545.
3. Pfann W G. Zone Melting .New York: Wiely ,1996
4. Zhou Raohe, Solidification Technology, [M]. Beijing:
5. Zhang kecong. Crystal Growth of Mechanics and Technology, Beijing: Science Press,1997.167.
6. Huhanqi.The principle of metal solidification[M]. Beijing : Beijing University of Science and Technology,2000 : 221-229, 109-151.

Ex. 1016, Page 1075

ECS Transactions, 18 (1) 1049-1058 (2009)
10.1149/1.3096571 © The Electrochemical Society

Performance Limitations of mc-Si Solar Cells Caused by Defect Clusters

Bhushan Sopori,[1] P. Rupnowski,[1] V. Mehta,[1] V. Budhraja,[1] S. Johnston,[1] N. Call,[1] H. Mountinho,[1] M. Al-Jassim,[1] Aziz Shaikh,[2] M. Seacrist,[3] and D. Carlson[4]

[1]National Renewable Energy Laboratory, [2]Ferro Electronic Materials, [3]MEMC Electronic Materials, [4]BP Solar

Multicrystalline silicon wafers used for solar cells exhibit defect clusters—localized crystal defects in and near grains of some specific orientations. Defect clusters are also dominant sites for impurity precipitation, and they remain ungettered and unpassivated through the solar cell processing. This paper describes characteristics of defect clusters, and shows, through theory and experiment, that defect clusters typically lower cell efficiency by 3 to 4 absolute percentage points. To recover this efficiency loss, it is necessary to getter precipitated impurities.

Introduction

Multicrystalline silicon (mc-Si) constitutes more that 50% of the total Si for photovoltaic (PV) applications and this share is expected to grow rapidly. The techniques for growing mc-Si by casting or by shaped crystal growth have been developed in the last 30 years primarily for solar cell applications to lower the wafer cost. Multicrystalline Si does not require seeding, and furnace designs are simpler that Czochralski (CZ) or float-zone (FZ) methods; therefore, the cost of casting and ribbon growth techniques is considerably lower. Also, mc-Si is produced from a lower-grade feedstock, consisting primarily of reject and out-of-spec material from the microelectronics industry. Consequently, the mc-Si wafers contain high concentrations of impurities. The lower material purity does not impact the quality of the ingot to the same degree as in CZ and FZ methods. But, because mc-Si material is grown very rapidly, the ingot experiences higher thermal stresses, accompanied by high defect densities. Dislocations are the primary intragrain defects. The impurities interact with defects during crystal growth, which leads to a very strong influence of defects on the cell performance [1].

This paper presents theoretical and experimental results of our studies to determine the performance limitations set by defect clusters (DCs) in mc-Si solar cells. DCs in mc-Si wafers are manifested as spatially localized, large regions of high dislocation density, around grains of certain preferred orientations. The presence of DCs has a strong influence on the material properties and cell performance. The objective of our comprehensive study on DCs is to understand their nature and their influence on cell performance, and to assess the performance improvement that can be realized if the influence of DCs can be eliminated. The effect of grain orientation on the defect density is also investigated. Finally, we explore possibilities of eliminating the formation of defect clusters during crystal growth or treating them during cell processing to minimize their electronic influence.

Characterization of Defect Clusters

As-grown mc-Si wafers have high concentrations of C and/or O. They also contain transition-metal impurities (e.g., Fe, Cr) in levels reaching 10^{14} cm^{-3}, which are detrimental to the minority-carrier lifetime. Typically, the average minority-carrier lifetime of as-grown material is < 10 μs. The average defect density of current mc-Si wafers is about $5x10^5$ cm^{-2}. But, a unique feature of current mc-Si wafers is that they contain "defect clusters"—crystal defects that clump together, forming extended defected regions that remain separated from each other.

A detailed structure of a defect cluster can be delineated after wafer polishing and defect etching [2]. Figure 1a is an optical micrograph of a defect-etched, cast mc-Si wafer showing dislocations piling up in a defect cluster that extends over several grains. Similar clustering is seen in ribbon material. Typically, the fractional area of the wafer covered by defect clusters is about 5%–10% for a high-quality material; lower-quality material may have a higher-percentage area covered by defect clusters. On a larger scale, clustering of defects can also be observed on defect maps generated by the commercial instrument PVSCAN. An example of such a map, acquired for a commercial 4.25-in x 4.25-in mc-Si wafer, is shown in Fig. 1b. This map reveals that the majority of the wafer has a low (and much of it nearly zero) dislocation density [3]. The average value of the dislocation density is about 4 x 10^5 cm^{-2}. The presence of defect clusters can be seen as light regions in this otherwise very low defect density material. These regions can have defect densities as high as 10^7 cm^{-2}.

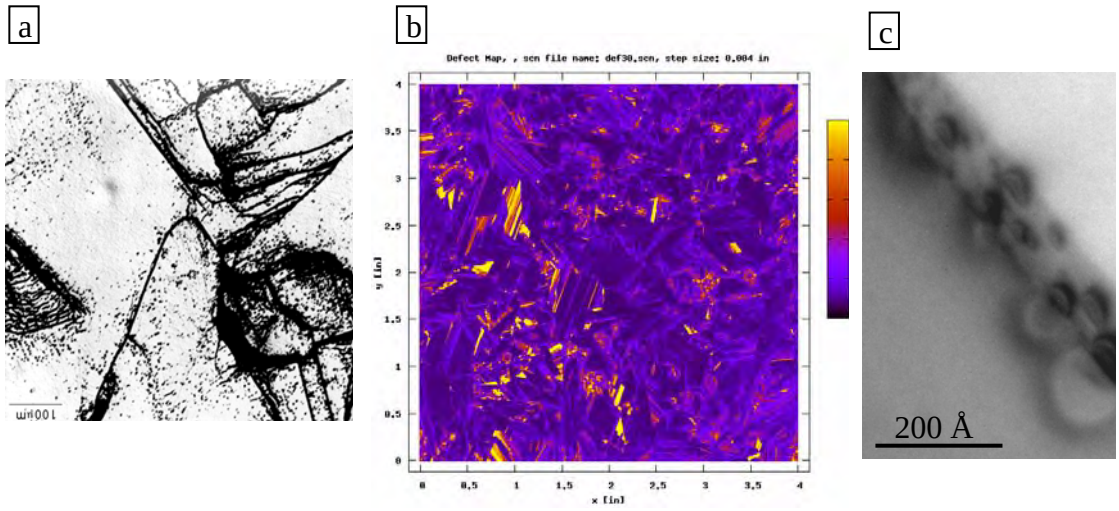


Figure 1. (a) Optical micrograph of a DC in a defect-etched wafer, (b) PVSCAN map showing density of defects on 4-in. wafer, and (c) TEM image of impurities precipitated in a DC.

Defect clustering occurs during crystal growth when local thermal stress exceeds yield stress of some preferred grain orientations, causing the stress relief through local generation of defect networks. Defect clusters also serve as internal gettering sites for metallic impurities, and they often result in impurity precipitation at these sites [4]. Figure 1c depicts a cross-sectional transmission electron microscopy (TEM) image of such precipitates in a DC region; other researchers also observed similar precipitates. Precipitation of impurities at defect clusters is of particular concern, because it is now

Ex. 1016, Page 1078

known that precipitated impurities are difficult to getter by the techniques used in the PV industry (namely, P diffusion for junction formation and Al alloying for back-contact formation). Hence, defect clusters have an important bearing on the development of new impurity-gettering methods that must dissolve the precipitates during gettering.

Formation of Defects

In a mc-Si casting or ribbon, the grains with orientations of lower yield stress respond to the thermal stress by generating dislocations. As the dislocations are generated, the stress is relieved. This allows other grains to escape dislocation generation. To study the correlation between crystal orientation and dislocation density, we had to acquire grain-orientation maps. We used a laser scanner to map grain orientations (Fig. 2) for wafers originating from various sections of an ingot. After calibration, such maps are compared with defect density images. Our initial study suggests that dislocations are generated in grains represented by blue and violet orientations [5].

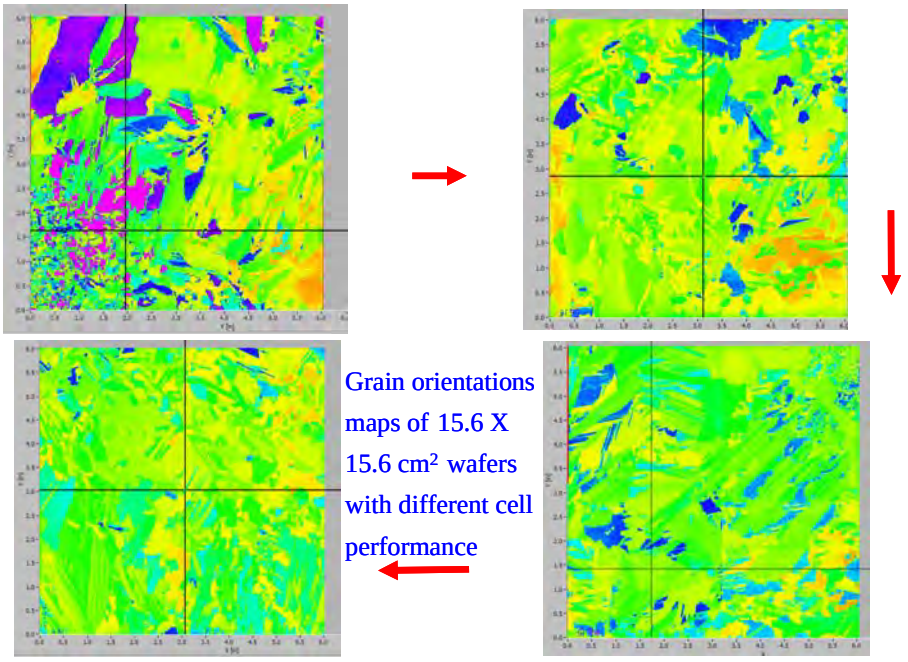


Figure 2. Orientation distribution maps of 6" wafers taken from different sections of an ingot. Each color represents a specific orientation.

Electro-Optical Response of A Cell with Defect Clusters

Modeling

The influence of defect clusters on the electro-optical response of a cell can be evaluated using a network model. In this model, we consider a defect cluster as a localized, large defect that propagates through the entire cell (crossing both the base and emitter regions of the cell), as illustrated in Fig. 3a [6]. Because of very high recombination and large size, one can ignore internal carrier transport and band bending associated with each

defect cluster. The defected region acts as a "poor" device in the spatial distribution of the total cell.

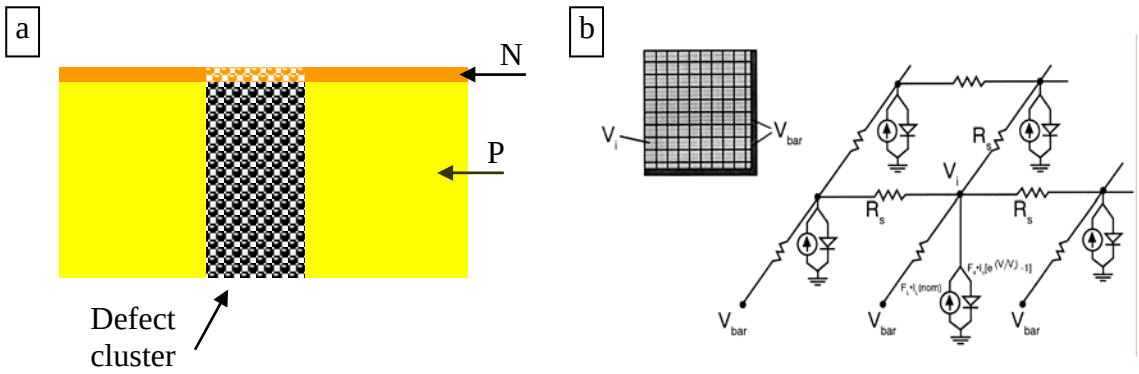


Figure 3. (a) Schematic of a defect cluster, and (b) a network model of a solar cell showing voltage and current sources corresponding to dark (with subscript d) and illuminated (with subscript L) conditions, and the resistive components due to the sheet rho of the junction.

The modeling requires two steps. First, each device is represented in terms of the recombination properties associated with its defect density, which yields values of photo-generated current density (J_{ph}), minority-carrier lifetime (τ), and dark saturation-current components J_{01} and J_{02}, corresponding to the bulk and the junction recombination, respectively. Next, the diode array is interconnected using resistive components corresponding to the sheet resistivity of the junction and the metallization pattern. The network is solved to yield the terminal characteristics of the device, as well as distribution of local currents and voltages for any given terminal voltage.

In the network model for defect clusters, we basically have two kinds of diodes: a defect-free diode and a defected diode. The characteristics of each diode can be expressed in terms of the photocurrent, J_{ph}, and two exponential components of the dark current, J_{dark}, in a standard form as [7,8]:

$$J_{dark\ i}(V) = J_{01i}.\exp\{(-eV/kT) - 1\} + J_{02i}.\{\exp(-eV/2kT) - 1\}, \quad (1)$$

where *i* denotes a diode index and J_{01}.and J_{02} represent the saturation currents. The total current through a diode, J_i, is then given by:

$$J_i = J_{ph\ i} - J_{dark\ i}(V) . \quad (2)$$

The values of $J_{ph\ i}$, J_{01i}, and J_{02i} can be estimated from the experimental measurements. For example, we select one cell and make an estimate of J_{ph} values for defect-free and defected cells based on laser-beam-induced current (LBIC), or long wavelength, responses and cell current-voltage (I-V) plots. However, J_{01} and J_{02} cannot be determined from the cell itself. We have built a library of J_{01} and J_{02} values for a variety of different materials and for different defect densities, using a diode array technique described in Ref. [9]. Edge-passivated, mesa diode arrays are fabricated on wafers and their electrical characteristics are probed. The device characteristics and their defect data are compiled

Ex. 1016, Page 1080

and used as input in our model. These data result in excellent agreement between calculated and actual terminal characteristics of the large-area cell (as seen in next section). It should be pointed out that the network model we have developed assumes no internal carrier transport—the communication between the devices occurs via a highly conducting emitter region and the bus bar.

Solar cells with measured dislocation distributions were simulated using our network model. Typically, 3000 individual diodes were used to represent a single large-area solar cell. Figures 4 and 5 show simulation results that were performed for a material provided by a commercial vender. Calculated I-V characterisitics of two defected cells and one without defects are compared in Fig. 4. As seen, defect clusters have a dominant effect in degrading the voltage-related parameters.

In Fig. 5, open-circuit voltage (V_{oc}) and short-circuit current density (J_{sc}) are presented as a function of the relative area covered by defect clusters. For comparison, measured values of V_{oc} and J_{sc} are also included. The parameters of an ideal solar cell, which does not exhibit any active defect clusters, were extrapolated from the obtained data points of Fig. 5. We found that the extrapolated ultimate values for an ideal cell were V_{oc} = 615–620 mV and J_{sc} = 31 mA/cm^2. It is important to point out that these values depend strongly on the initial material properties and cell processing.

Our network model provided detailed information on the spatial distribution of voltage and current over the entire cell for a given terminal voltage. Figure 6 depicts results of our simulation for a 4-in. x 4-in. cell whose LBIC map (at λ = 0.98 μm) is shown in Fig. 6a (identifying defect clusters). Figures 6b, 6c, 6d, and 6e present spatial distribution of the direction-of-current flow at forward terminal voltages of 0.500, 0.505, 0.515, and 0.525 V, respectively. The green color indicates a forward current and red color indicated a reversal in the current flow. Thus, defect clusters can cause a local current to flow in the opposite direction, causing net internal energy dissipation at the DC sites.

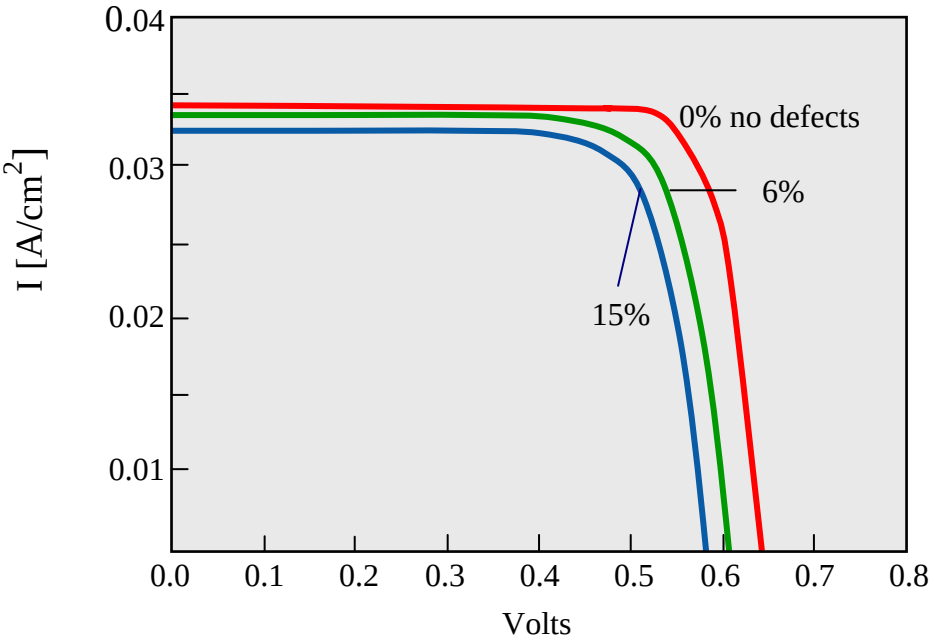


Figure 4. A comparison of calculated I-V characteristics of three cells with 0%, 6%, and 15% of area covered by defects.

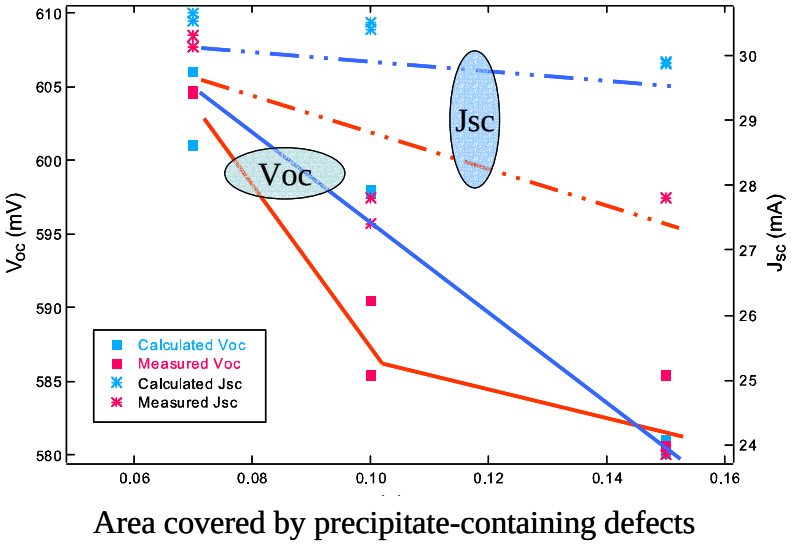


Figure 5. A comparison of theoretical V_{oc} and J_{sc} and experimental data on 134-cm^2 cells.

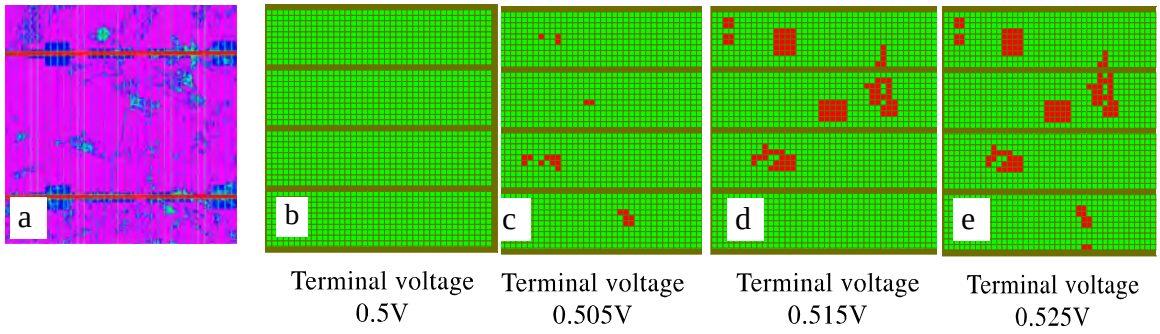


Figure 6. (a) **LBIC map (at λ = 0.98 μm) of a 4-in. cell, showing the bulk response of the map and the presence of defect clusters. The main defect clusters are labeled. (b–e) Spatial distribution of the direction of current flow at various forward terminal voltages. Green color indicates a forward current and red color indicates a reversal in the current flow.**

Experimental Studies

To measure the effect of defect clusters on cell response, the following experimental procedure was used. Sequential mc-Si wafers were obtained from different vendors, and alternate wafers were processed into cells. The unprocessed wafers were used to obtain dislocation/grain boundary maps by PVSCAN, and grain orientation maps. This information was used to determine average area coverage by DCs and to establish a relationship between DC locations and grain orientations. The processed solar cells were characterized by dark and illuminated I-V plots, and by photoresponse mapping. These results confiremd that defect clusters can produce shunts in the device. To further investigate shunting mechanisms, we performed thermal and electroluminescent imaging of the cells and compared them with defect and LBIC maps. Some samples were used to

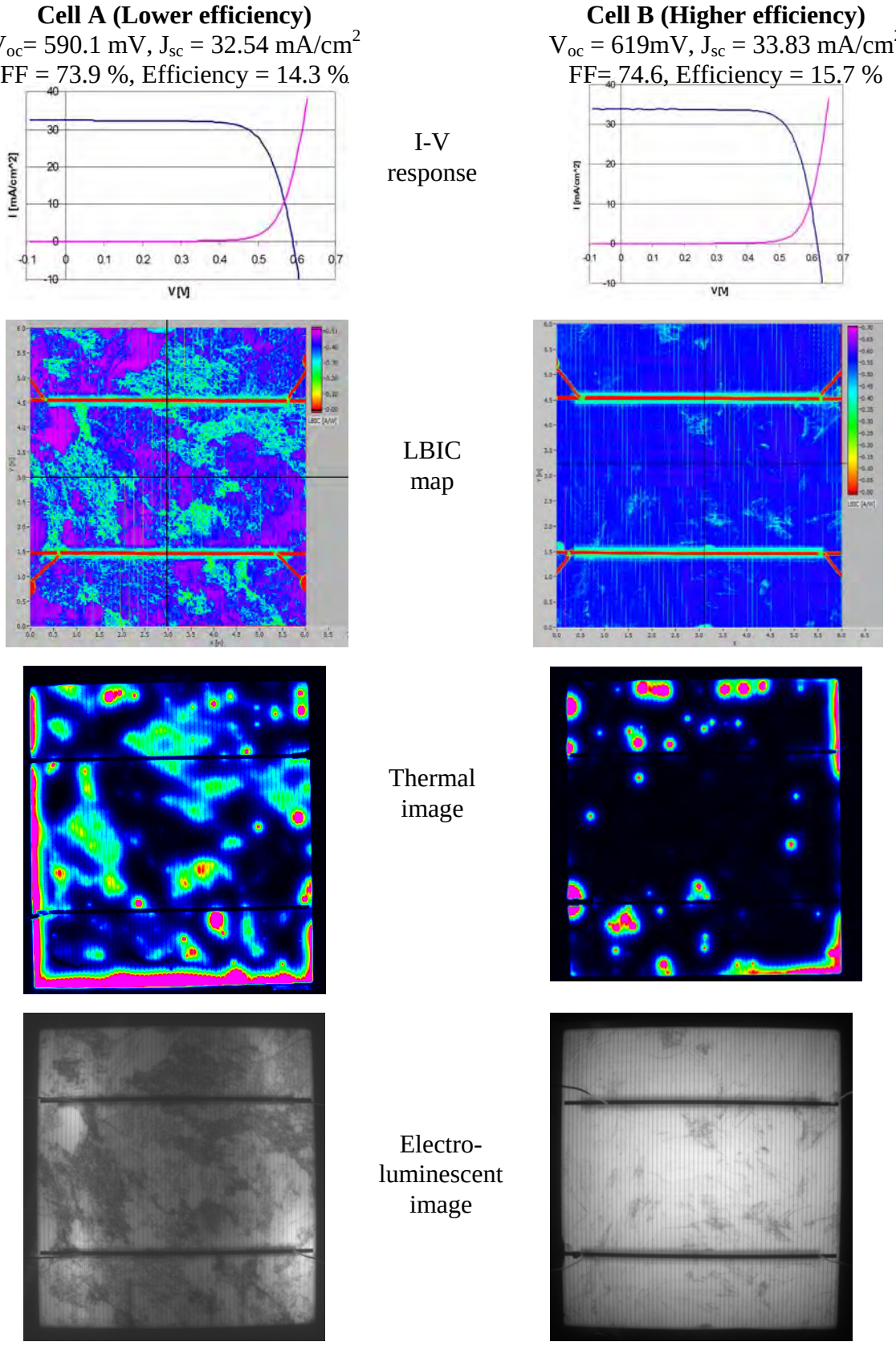


Figure 7. I-V response, 980-nm LBIC, thermal, and electroluminescent image of 6-in. x 6-in. cells. Two cells of lower efficiency (left column) and higher efficiency (right column) are compared.

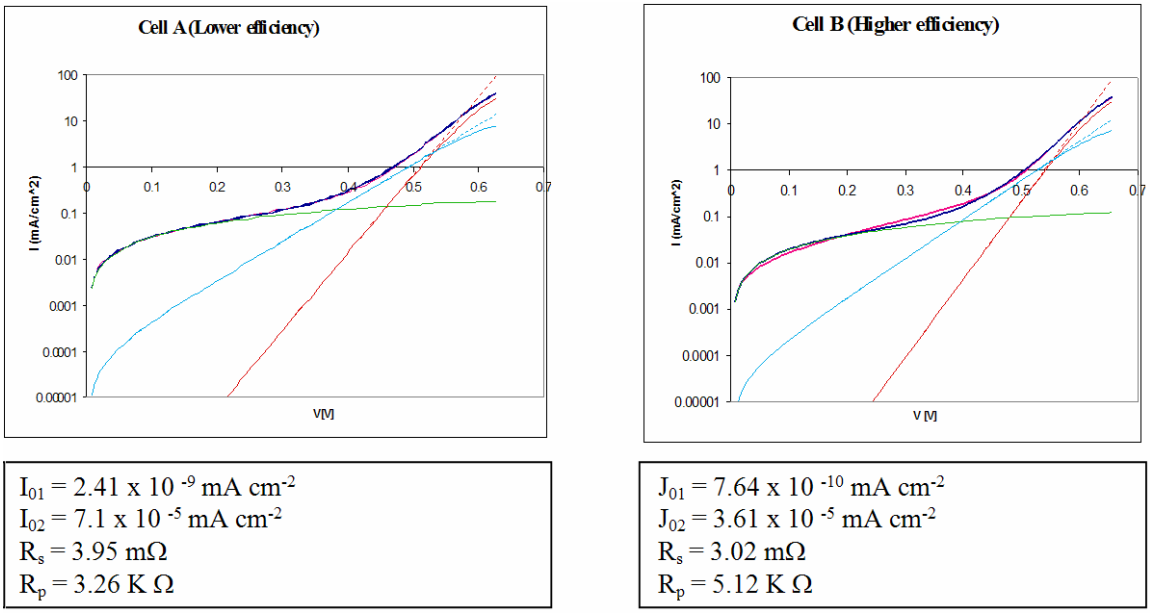
perform cross-sectional analyses by TEM, atomic force microscopy (AFM), and capacitive atomic force microscopy (CFM).

A comparison of two 6-in. cells labeled A and B is presented in Figs. 7 and 8. These two cells were chosen because they belong to two different efficiency categories. The efficiency of cells A and B are 14.3% and 15.7%, respectively. Figure 7 includes light and dark I-V response, basic cell parameters, LBIC maps for wavelength of 980 nm, as well as thermal and electroluminescent images. It can be observed that all three cell parameters—V_{oc}, J_{sc}, and fill factor (FF)—are lower in the case of cell A. Further investigation of LBIC maps clearly revealed the structural difference between the cells. Cell B yielded a very uniform response over the entire area, with only a few spots with active defects. Contrary to cell B, cell A contains extensive active defect clusters that reveal themselves as regions of lower response. In general, thermal images show shunting areas where the energy radiates outward in the form of heat. For cell A, much heat is generated in various areas of the cell. The locations of these hot spots overlap with areas of low LBIC response. Therefore, it is clear that active defect clusters are responsible for poor performance of this cell. For higher-efficiency cell B, there are much fewer hot spots. Comparing the figures, it can also be observed that the hot spots and active defect clusters are not coincident. This means that some other mechanisms dominate in shunting generation in cell B.

Electroluminescent maps in Fig. 7 clearly revealed the material's nonuniformities. The location and size of active defect clusters can easily be determined. A very good correlation between LBIC and electroluminescent maps was found for cells with both good and bad performance.

Another useful and complementary approach to investigate the influence of DCs is to examine the electronic characteristics of the entire cell by dissecting the dark and illuminated I-V plots. Here, we show some results of deriving important parameters that characterize the behavior of a cell by fitting the experimentally measured dark I-V plots with the theoretical two-diode model (i.e., applying Eq.[1] to the entire cell). It is important to recognize that we are obtaining lumped parameters of a distributed device. This procedure yields J_{01}, J_{02}, series resistance, and shunt resistance of each device. The quality of the fit can be determined by regenerating a theoretical I-V plot and comparing it with the experimental plot. Figures 8a and 8b show the values of the parameter values determined by this procedure for two cells described in Fig. 7. Figure 8 also shows that the theoretical and experimental I-V plots have excellent agreement. One can see that the higher-efficiency device, which contains a lower density of DCs, has lower values of J_{01} and J_{02}, a lower series resistance, and higher shunt resistance.

Ex. 1016, Page 1084



I_{01} = 2.41 x 10 $^{-9}$ mA cm^{-2}
I_{02} = 7.1 x 10 $^{-5}$ mA cm^{-2}
R_s = 3.95 mΩ
R_p = 3.26 K Ω

J_{01} = 7.64 x 10 $^{-10}$ mA cm^{-2}
J_{02} = 3.61 x 10 $^{-5}$ mA cm^{-2}
R_s = 3.02 mΩ
R_p = 5.12 K Ω

Figure 8. Illustration of fitting measured I-V characteristics to determine J_{01}, J_{02}, R_s (series resistance), and R_p (shunt resistance).

Conclusions

This paper presents a combination of numerical and experimental methods that were used to characterize defect clusters in mc-Si solar cells. In particular, we described a defect-generation mechanism, grain-orientation effect, interaction of defects with impurities, effect of defect distribution on cell performance, and shunting mechanisms associated with defect clusters.

The dependence of dislocation density on grain orientation is very interesting and lends itself to a well-known observation: cell performance depends on the location of the wafer within the ingot. A comparison of the orientation and dislocation maps shows that dislocation distribution mimics the orientation distribution. This result has very important implications on potential methods to minimize the formation of dislocation clusters.

Experimental and numerical results both show that defect clusters are regions of low minority-carrier diffusion length and local shunts that cause dissipation of energy internally within the cell. There is good agreement between theoretically predicted parameters of cells with defects and corresponding experimental data. We also found that the best solar cells from commercial vendors performed nearly as well as predicted by the theoretical model that does not have any active defects. The best cell performance was limited by the impurities in the material.

The salient results of our analysis are the following: (i) Defect clusters are the primary sites of impurity precipitation, and hence, are not gettered by conventional gettering treatments of P diffusion and Al alloying. (ii) Defect clusters act as shunts, creating "internal sinks" that dissipate power within the cell. (iii) The loss depends on several factors—the nature of the impurity precipitates, distribution and total area of defects

Ex. 1016, Page 1085

compared to the cell area, quality of the base material (the regions with no defects, in which cell performance is limited by the dissolved impurities only), and cell processing technology. (vi) In a typical cell, the defect clusters produce an efficiency loss of 3 to 4 absolute percentage points in efficiency.

To recover the efficiency loss caused by cluster defects, it is necessary to either eliminate them during crystal growth or getter precipitated impurities at DC sites. The information gathered on defect clusters suggests potential techniques that can minimize defect cluster formation and/or dissolve precipitated impurities at temperatures and times that are compatible with solar cell fabrication; these techniques offer the potential to raise efficiency of mc-Si solar cells to >20%. An intuitive approach to mitigating the influence of defect clusters would be to dissolve the precipitates by a high-temperature treatment prior to or during gettering. A new approach, recently developed at the National Renewable Energy Laboratory, dissolves impurity precipitates by using vacancy injection to lower the temperature needed for precipitate dissolution.

*Employees of the Alliance for Sustainable Energy, LLC, under Contract No. DE-AC36-08G028308 with the US. Dept. of Energy have authored this work. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-lip, irrevocable, worldwide license to publish or reproduce the published form of this work. or allow others to do so, for United States Government purposes.

References

1. B.L. Sopori, *Proc. ICDS-19, Trans Tech Pub.,* edited by G. Davies and M.H. Nazare, 527 (1997).
2. B.L. Sopori, *J. Electrochem. Soc.* **131**, 667 (1984).
3. PVSCAN uses optical scattering from a defect-etched wafer to statistically count defects. It is also fitted with capabilities to make LBIC and reflectance measurements at 0.63 and 0.98 μm. It is available from GT Solar Technologies, Nashua, NH 03054. Also see B.L. Sopori, U.S. Patent 5,581,346.
4. T. Buonassisi, O.F. Vyvenko, A.A. Istratov, E.R. Weber, G. Hahn, D. Sontag, J.P. Rakotoniaina, O. Breitenstein, J. Isenberg, and R. Schindler, *J. Appl. Phys.* **95**, 1556 (2004).
5. B. Sopori, P. Rupnowski, D. Balzar, P. Sheldon, "Dislocation Generation in Si: A Thermo-Mechanical Model Based on Measurable Parameters", *Conference Record of the 2006 IEEE 4th World Conference on Photovoltaic Energy Conversion*, May 7-12, 2006, 936 - 939
6. B.L. Sopori and R.A. Murphy, *Proceedings of 12th European Photovoltaic Solar Energy Conference*, April 1994, Amsterdam, The Netherlands, 1797.
7. J.G. Fossum and F.A. Lindholm., *IEEE Trans. ED-27* **692**(1980).
8. B.L. Sopori, R.A. Murphy, and C. Marshall, *Proceedings of the 23rd IEEE Photovoltaic Specialists Conference*, Louisville, Kentucky, May 10–14, 1993, 190.
9. B.L. Sopori, *Appl. Phys. Lett.* **52**, 1718 (1988).

ECS Transactions, 18 (1) 1059-1064 (2009)
10.1149/1.3096572 © The Electrochemical Society

CLUSTER ION IMPLANTATION SYSTEM: CLARIS FOR BEYOND 45NM DEVICE FABRICATION (II)

Masayasu Tanjyo, Nariaki Hamamoto, Tsutomu Nagayama, Sei Umisedo, Yuji Koga, Noriaki Maehara, Hideyasu Une, Takao Matsumoto, Nobuo Nagai, and John O. Borland [a]

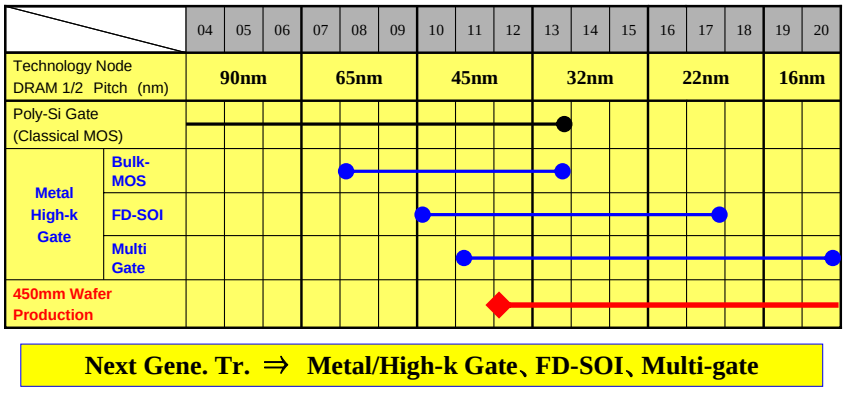
Nissin Ion Equipment Co.,ltd., Kuze-tonoshiro-cho, Minami-ku, Kyoto 601-8205, Japan
[a]JOB Technology, 98-1204 Kuawa St. Aiea, Hl 96701, USA

Newly developed sweep beam Cluster ion implanter: CLARIS with 0.2-7keV energy range for Boron beam and 1-10keV energy range for Carbon beam is introduced. Novel Cluster ion implantation technology is capable for 45nm beyond device requiring USJ formation (<15nm) with high retain dose (>70%) and low sheet resistivity (<1200Ω/sq). Comparison of retain dose and sheet resistivity of B_{18}, BF_2, and B beams with FLA shows the superiority of the B18 implantation for less than 500eV implantation.

INTRODUCTION

In Table 1, it is shown the transition of the transistor structure and material in ITRS2007 [1]. From 2008, the poly-Si gate of classical MOS is transferred to the high-k + metal-gate of new concept transistor. There are some kinds of new configuration devices, but the formation of the ultra-shallow junction (USJ) is continuously the key technology to fabricate next generation devices, as shown in Fig.1. Beyond 45nm node, the USJ for source drain extension is requiring low energy implantation below 0.5 keV. These low energy implantations have severe problems such as energy contamination with decel-mode operation and spot beam divergence which is caused from space charge effects. Solving these problems there are three kinds of approaches are known. One is the classical method to improve the deceleration lens, which is more reliable but less attractive. Second is the innovative method called plasma immersion implantation (PIII), which is applicable for low energy ultra-high dose like over 1E16/cm^2, but which is not so easy to answer the requirement of severe beam controllability. [2] Last one is the cluster ion implantation, which is expected to be precise dose, uniformity and beam angle control method. [3], [4], [5]
In this paper Nissin Cluster ion implanter system named CLARIS is introduced, and the characteristics of the Cluster ion implantation technology is explained.

Table 1. Transition of Transistor Structure in ITRS2007



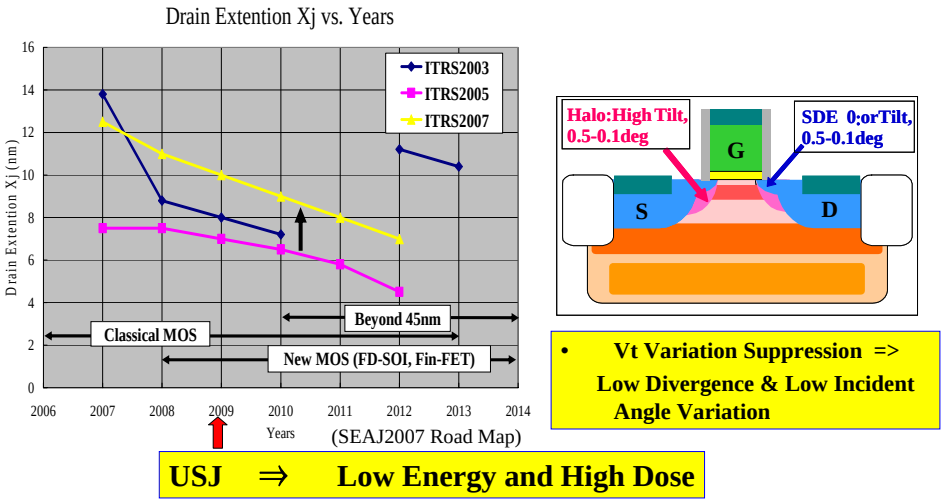


Figure 1. ITRS Road Map ; Transition of Junction Depth Xj

CLUSTER ION IMPLANTATATION SYSTEM: CLARIS

In Fig.2, it is shown the CLARIS layout. Its foot print is almost similar to our medium current ion implanter EXCEED3000AH and the end station is same to it. But, the ion source is developed for Cluster beam production and the beam line is originally designed for high mass number beam transportation.

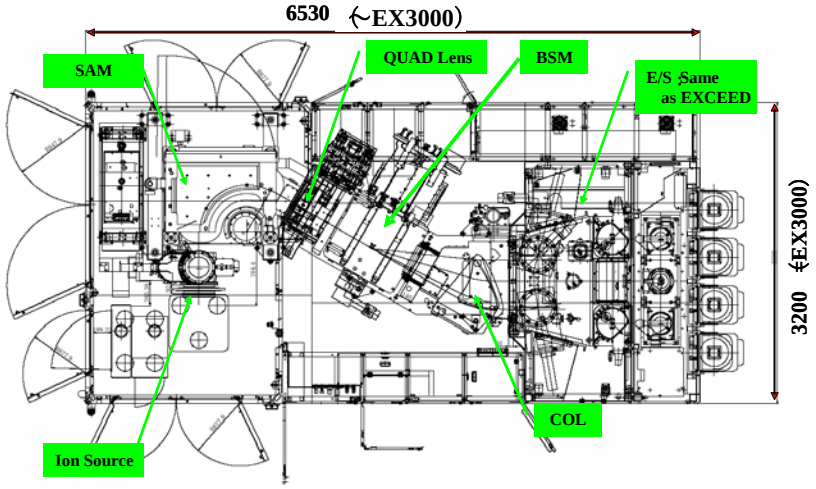


Figure2. CLARIS Layout

Equivalent Energy (keV)	Equivalent Sweep Beam Current (uA)	Ion Species
0.2	1,800	$B_{18}Hx^+$
0.3	4,000	
0.5	8,000	
1	14,400	
3	20,700	
5	15,000	$B_{10}Hx^+$
6	15,000	
7	15,000	

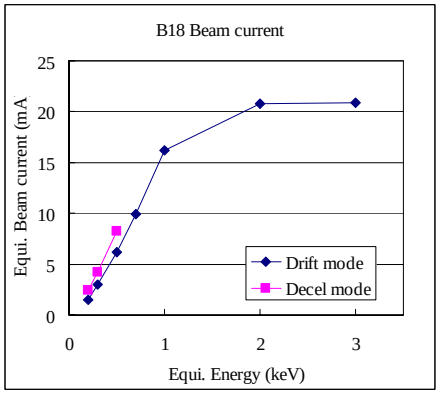


Figure 3. Boron Beam Current Specification

In Fig.3, it shows the Boron beam current specification. For 0.3keV Boron beam of 4mA makes 32 WPH at 1E15/cm^2, which is very difficult with conventional Boron monomer implantation without any energy contamination.
In Fig.4, it shows the Carbon beam current characteristics. For Carbon Cluster beam, $C_{14}H_{14}$ or $C_{16}H_{10}$ are used as source materials. As the $C_{14}H_{14}$ is easily resolved to C_7Hx^+ ions that it is suitable for stress engineering of deep implantation.[6] Contrary, as the $C_{16}H_{10}$ is capable for low energy high beam current, that it is used for Carbon co-implantation with Boron Cluster for suppress the Boron diffusion at annealing. [7]

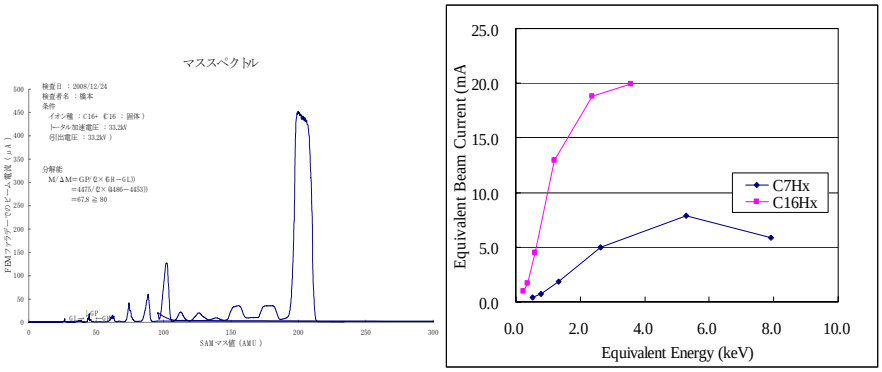


Figure 4. Carbon Beam Current Characteristics

In Table 2, it is summarized CLARIS G1 specification.

Table 2. CLARIS G1 Specification

Tool Size	3200(W)*6530(L) *3200(H)
Ion Source	EB source
B/L	Fast magnetic beam scanner & Collimator Hybrid implantation
E/S	Same as EXCEED series
Energy	$B_{18}H_{22}$; 4~60KeV (B equivalent: 200eV~3keV) $B_{10}H_{14}$*; ~80KeV (B equivalent: ~7keV)
Dose range	9E12~3.6E16 atoms/cm2
Equivalent Energy & Beam Current	$B_{18}H_{22}$ 200eV : 1.8mA 300eV : 4mA 500eV : 8mA 1 keV : 14.4mA 3keV : 20.7mA $B_{10}H_{14}$* 5keV : 15mA 7keV : 15mA
Uniformity Repeatability	<1% (Boron Energy>= 1keV) <1.5% (Boron Energy < 1keV)
Horizontal Parallelism	< +- 0.5 degree
Metal contamination	Al < 50ppm, others < 10ppm
Particle	<30pc (particle size > 0.12um)

* $B_{10}H_{14}$: Option Preliminary Specifications. Subject to Change without Notice.

CLUSTER ION IMPLANTATION TECNOLOGY

In Table 3, it shows the next generation beyond 45nm transistor requirement and the solution of the Cluster ion beam technology.
For the productivity improvement, cluster implantation is capable for the effective energy reduction and the enhance beam current with the molecular atom numbers.
The space charge effect reduction makes low beam divergence and low wafer charging, which will suppress the Vt variation. For example, using Octa-decaborane ($B_{18}H_{22}$), the space charge effect is reduced to $1/N^2$ =324 (N=18).

Table 3. Next Generation Transistor Requirement & Cluster Technology

No.	Requirement	Measure	Method
1	USJ	Low-E High-Current	B_{18},As_4
2	Suppression for Vt variation	Low-div. angle, low injection angle variation	Low space charge
3	Low thermal budget	High-activation	Self-amorphization &re-crystallization
4	Increase on-current	Stress engineering, Multi-gate	C_7H_7-Si:C、High angle I/I

In Fig.5, it shows TW and Rs uniformity at B, 0.6keV and SIMS depth profile at B, 0.3keV, respectively. It is recognized that at these low energy condition the uniformity and energy contamination are no problem.

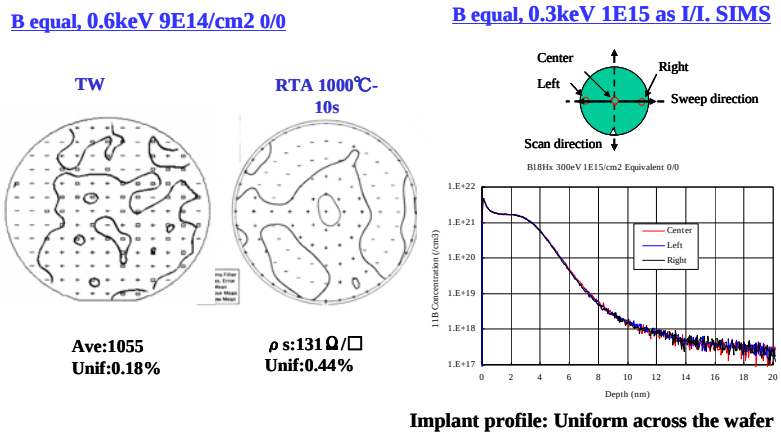


Figure 5. TW/Rs Uniformity & SIMS Profile

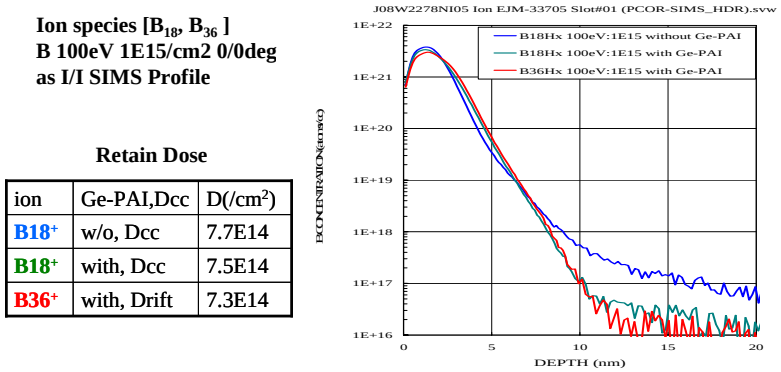


Figure 6. Ultra-low Energy B 100eV: $B_{18}Hx^+$ & $B_{36}Hx^+$ SIMS Profile

In Fig.6, it shows the SIMS profile of $B_{18}Hx^+$ and $B_{36}Hx^+$ implantation at B, 100eV energy. With Ge-PAI condition it shows no energy contamination that the B_{18}^+ w/o Ge-PAI profile shows channeling effect, which is almost less than 1E18/cm^3 level. It also shows the retain doses by integrating the B profiles, which are around 70 – 80 % with tilt 0deg twist 0deg implantation condition.

For the evaluation of the retain dose and the sheet resistivity characteristics, it was done the comparison of the $B_{18}Hx^+$, BF_2^+, B^+ implantation with flash lamp anneal (FLA). Wafer tilt angle is also changed from 0 to 60 degree. Figure 7 shows the Rs and TW of tilt angle dependency. At this evaluation, beam energy is controlled for same projection depth that is for 0deg-tilt with 0.4keV and 60deg-tilt with 0.8keV. [8]
It is shown that the tilt angle dependency is almost flat except BF_2^+ which is machine energy resolution problem, that Rs is almost follows to cosine law. As the B+ energy is slight higher, that values of Rs are $B_{18}Hx^+ < B^+ < BF_2^+$, contrary values of TW are $B_{18}Hx^+ < BF_2^+ < B^+$. For $B_{18}Hx^{+,\ values}$ of Rs and TW are both almost half to B^+ and BF_2^+.

.

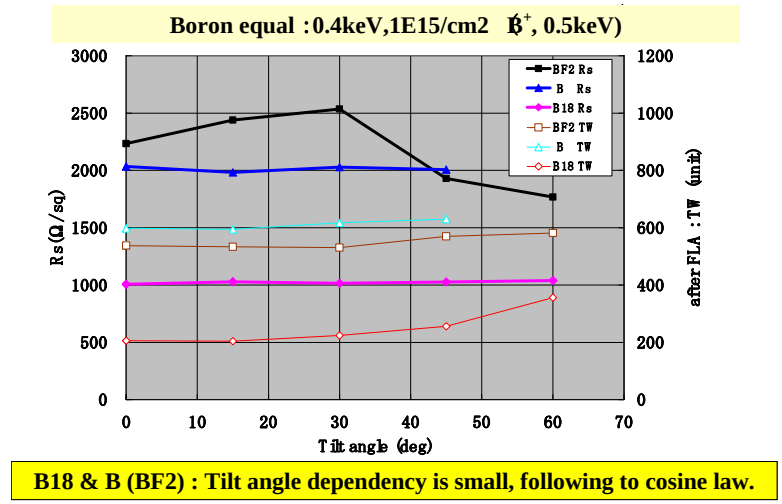


Figure 7. After FLA, Rs & TW Tilt Angle Dependency; B_{18}^+, BF_2^+, B^+

In Fig.8, it shows (a)as I/I and (b)after FLA SIMS profiles with different ion spices and tilt angles. In both (a) and (b) cases, B_{18}^+ shows most abrupt profile and shallow junction.

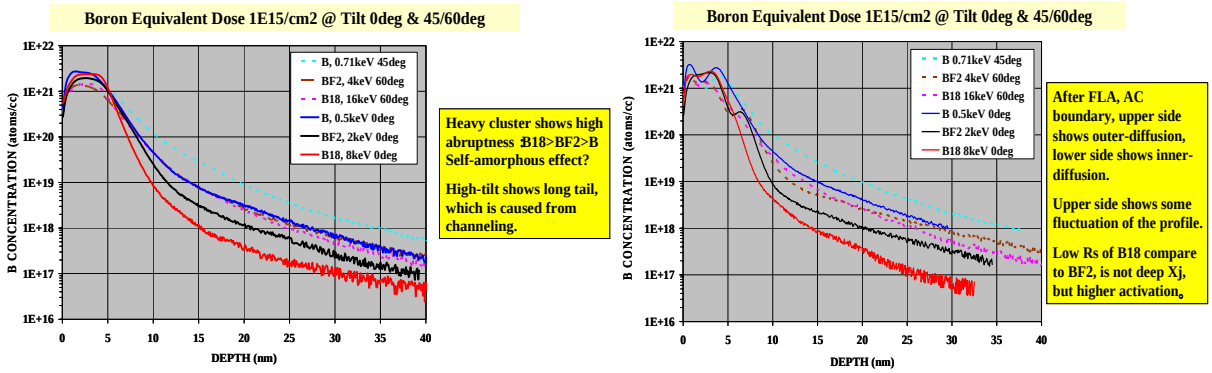


Figure 8. (a) as I/I & (b) after FLA SIMS Profile; B_{18}^+, BF_2^+, B^+

In Fig. 9, it shows (a) tilt angle and (b) energy dependency at high tilt of retain dose and Rs with different ion spices and tilt angles.
In Fig.9(a), values of retain dose of as I/I are $BF_2^+ < B_{18}^+ < B^+$, and values of retain dose of after FLA are $BF_2^+ \sim B_{18}^+ < B^+$. That is, B_{18}^+ is smaller loss for implantation but larger loss for annealing than those of BF_2^+. It is supposed that the shallower profile makes larger out diffusion at annealing. For high tilt condition the retain dose decreases 20-30%.
In Fig.9(b), values of retain dose of lower energy at 0.4keV are $BF_2^+ < B_{18}^+ < B^+$, and values of retain dose of higher energy at 1.6keV are $BF_2^+ \sim B_{18}^+ < B^+$. Values of Rs are also abruptly increases for low energy BF_2^+. That is, the retain dose of B_{18}^+ and BF_2^+ are

almost same but dopant activation ratio should be much difference. And as the energy is decreased the loss of BF_2^+ is enhanced and Rs is decreasing than that of B_{18}^+.
For B^+, it shows minimum loss for both implantation and annealing but activation ratio is least of the three, because of the deepest dopant profile.

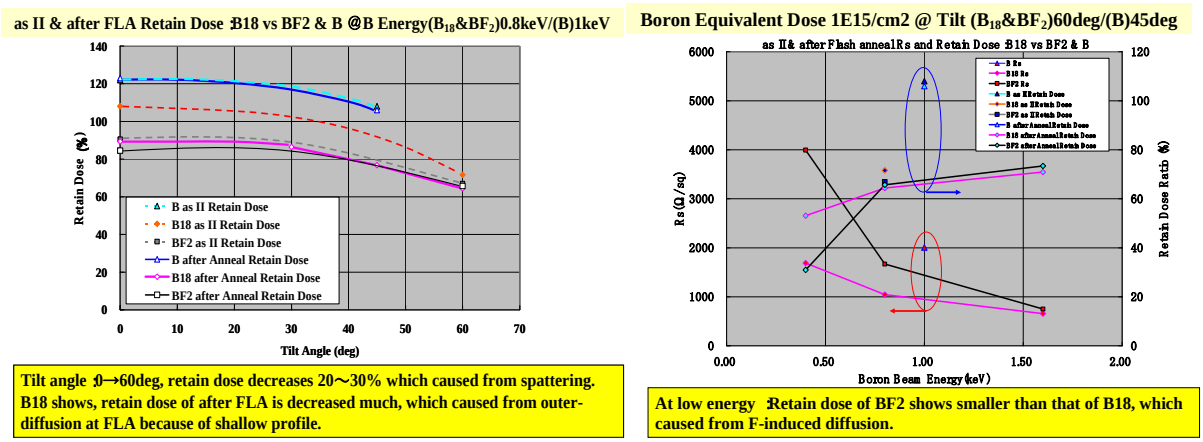


Figure 9. (a) Tilt Angle & (b) Energy Dependency of Retain Dose and Rs; B_{18}^+, BF_2^+, B^+

SUMMARY

1. Nissin Cluster Ion Implantation System: CLARIS for beyond 45nm node transistor mass-production is developed and Boron version: G1 is released.
2. Practical use of Carbon Cluster beam is under developing, and $C_7H_x^+$ and $C_{16}H_x^+$ ion beam current characteristics are evaluated. It is applied to co-implantation of P-and N-MOSFET and Stress engineering of N-MOSFET.
3. Boron Cluster beam shows superiors for retain dose that even at B 100eV 0deg-tilt condition, >70% retain dose is founded.
4. Comparison of B_{18}^+, BF_2^+, B^+ beams, it is confirmed that B_{18}^+ beam is capable of low sheet resistivity, which is caused from high activation than BF_2^+, B^+ beams.
5. For beyond 45nm node transistor mass-production, it is required the break-through technologies, that CLARIS is a candidate of those technologies.

Acknowledgments

This work is supported by Japan Science and Technology Agency (JST) for the equipment development and The New Energy and Industrial Technology Development Organization (NEDO) of Japan for the process development.
Elemental development of CLARIS is ought to SemEquip Inc, which is highly appriciated.

References

1. ITRS2007
2. Y.Sasaki et. al, IEDM 08-917
3. M.Tanjyo et. al, 7th ECS-ISTC 2008, pp.504-510
4. S.Umisedo et. al, 17th IIT 2008, pp.296-299
5. T. Nagayama et. al, 17th IIT 2008, pp.434-437
6. S-M Koh et. al, IEEE ED Letter, Vol.28, No.12, Dec. 2008. p.1315
7. A. Falepin et. al. IEDM2008 (978-1-4244-1738-4/08/S25.00 c2008 IEEE)
8. M.Tanjyo et. al, 36th JSAP Cluster Seminar(ISBN:978-4-903968-38-4)pp.59-67

Ex. 1016, Page 1092

ECS Transactions, 18 (1) 1065-1070 (2009)
10.1149/1.3096573 © The Electrochemical Society

Silicon Etching Technology Using Inductively Coupled Plasma Suitable for Fabrication of Silicon Nanowires and Ring Resonators

Jingtao Zhou[1], Huajun Shen[1], Huihui Zhang[1], Xinyu Liu[1]

[1]Institute of Microelectronics of Chinese Academy of Sciences 3 Beitucheng West Road, Chaoyang District, Beijing, PR China

Silicon-on-insulator (SOI) is an attractive platform for single chip photonic integration. Silicon nanowires fabricated by SOI substrate are fundamental components such as ring resonators in the optoelectronics integrated circuits. High performance of these devices requires silicon nanowires have vertical and smooth sidewalls. In this paper, we report the fabrication of silicon nanowires with vertical and smooth sidewalls using $SF_6/C_2H_4/Ar$ based inductively-coupled-plasma (ICP) etching, Ring resonator with 200nm gap was fabricated. This technology is suitable for fabrication of planar photonic integrated components and circuits.

Introduction

The aim of integrated optics technology is to realize a variety of different optical components on a single chip. Silicon-on-insulator (SOI) is an attractive platform for single chip photonic integration. Silicon nanowires fabricated by SOI material are fundamental components in the optoelectronics integrated circuits, not only because they serve as "wires" to connect each device in the circuits but they are the basic structure of some function devices such as ring resonators. There are three requirements for silicon nanowires and ring resonators: firstly, single-mode propagation requires the waveguide has vertical sidewalls; secondly, low propagation loss requires the waveguide sidewalls have very low roughness; finally, some critical small size such as the gap between waveguide and the ring should be accurately fabricated.

Based on these demands, this study investigates the optimum conditions for the silicon etching in an inductively-coupled-plasma reactive ion etching (ICP-RIE) system (Corial 200IL). Its anisotropic etching capability is achieved by the alter flow rate of the etching gas (SF_6) and the passivation gas (C_2H_4), and gas (Ar) which is introduced to produce physical sputter. The etching result is affected by several parameters including the source power, bias power and gas flow rate.

In this paper, we report the fabrication of silicon nanowires with vertical and smooth sidewalls using a $SF_6/C_2H_4/Ar$ based ICP etching. Ring resonator with 200nm gap was fabricated. Strict and accurate control in fabrication process may improve the performance of the MMI splitter. Furthermore, it will help the MMI splitter extensively apply in ultra-compact integration of photonic integrated circuits.

Fabrication of silicon nanowires

The SOI wafer is provided by Soitec company, the thickness of silicon top layer is 0.26 μ m and the BOX layer(SiO_2) is 1 μ m. The fabrication technology has been borrowed from the silicon microelectronics industry. Fig.1 shows the fabricating process.

Ex. 1016, Page 1093

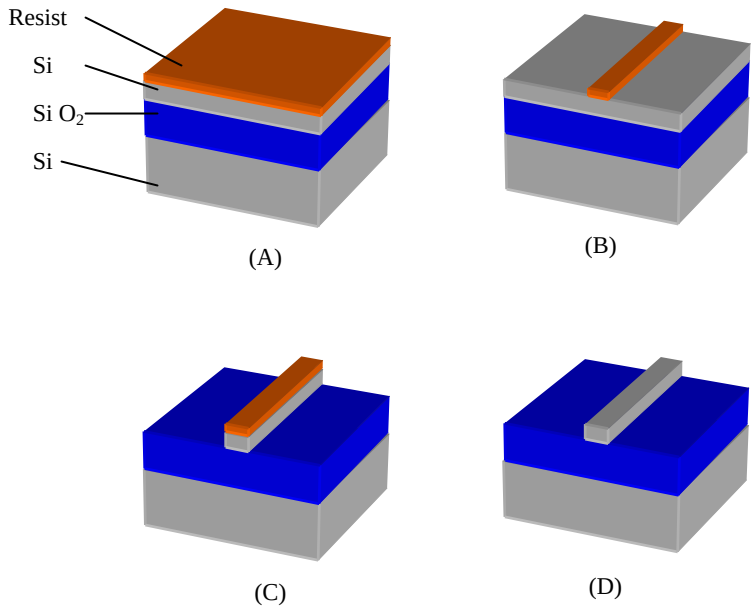


Figure 1.The fabrication process of silicon nanowires. There are total four steps included: (A) ZEP520 coating, (B) electron-beam lithography, (C) ICP etching, (D) resist removing. Among all these steps, the electron-beam lithography and ICP etching are the most essential.

Resist Coating

ZEP520 resist is spun onto the SOI wafer. After baking out the ZEP520 solvent, a resist film has formed and its thickness is about 300nm.

Electron Beam Lithography (EBL)

The second step is electron beam lithography (EBL), JBX6300FS EBL system has been used to generate the pattern of the silicon nanowires. The generation of the pattern by electron beam lithography is a very important step in the fabrication, which may highly affect the quality of the devices. The main parameters of the electron beam lithography include electron beam accelerating voltage, electron beam diaphragm diameter, exposal dose, electron beam scanning rate, electron beam pace, and operation distance. Besides, there are some other parameters such as diaphragm position, electric charge accumulation and so on [3]. It is necessary for us to optimize all the main parameters.

ICP Etching

The third step is ICP etching. Once written, the pattern is transferred into the semiconductor by dry etching. As mentioned above, high performance of silicon nanowires requires them have vertical and smooth sidewalls, so anisotropic etching is required. ICP technique is performed to etch the top silicon layer into dioxide layer of the SOI substrate. The fabrication parameters that are adjusted for an optimal result are also shown as follows: the gas resources are SF_6, C_2H_4 and Ar. We get a good result at

Ex. 1016, Page 1094

the condition that occurred with an RF power of 20W, LF power of 250W, pressure of 3mTorr, the optimize gas flux was SF_6 (sccm):C_2H_4: Ar = 3:1:50 or 4:2:50. The optimization of the etching conditions will be explained later.

Resist Removing

Finally, Resist should be removed by organic solvents.

Optimization of ICP Etching

The main factors which mostly influence the cross section shape of the etched silicon nanowires are the flux rates of the etching gases, the source power (LF power) and bias power (RF power). Through the optimization of the major factors, silicon nanowires with vertical and smooth sidewalls are achieved.

Optimization of gas flux rate

SF_6 and C_2H_4 are used as the etching and passivation gases respectively in this work. As we all know, SF_6 is a gas which can provide isotropic silicon etching [1]. In order to obtain vertical sidewalls, anisotropic silicon etching is particularly necessary. Then the passive gas C_2H_4 was introduced in the etching process, it can form a passivation polymer film on the surface of the sample including the exposed silicon area and waveguide sidewalls. The polymer film on the exposed silicon area can be removed by the reaction of the high energy plasma, and then the silicon can be etched deepen by plasma. On the other hand, the sidewalls are always protected from etching and thus the anisotropic etching is realized [2]. In order to have rapid etching rate, the Ar gas is introduced to produce physical sputter, thus fresh silicon surface can expose and continuously react with the plasma radical.

Table.1 lists the results of silicon etching under different gas flux rates. There is some regularity here: firstly, when only the etching gas SF_6 is added, silicon etching behavior exhibits isotropy. The reduction of the SF_6 gas flux may cause the decrease of silicon etching rate. For example, in Tab.1 the etching rate decreases from 750 to 450 nm/min when gas flux rate was changed from 40 to 10 sccm. Secondly, the passive gas C_2H_4 was added to obtain anisotropic silicon etching. Because it can form a polymer film on the surface of the sample, the increase of gas C_2H_4 flux may hinder silicon etching.

Comparing to the depth to be etched (about 260nm), the silicon etching rate is too high and vertical sidewalls are still unreachable this time. In order to obtain appropriate etching rate and vertical sidewalls, the gas flux rate of SF_6 may reduced to 3-4 sccm, at the same time Ar gas is introduced to produce physical sputter, so fresh silicon surface can expose and continuously react with the plasma radical. Finally, two optimized results which cause vertical and smooth sidewalls have been obtained. One condition is SF_6 (3sccm):C_2H_4 (1sccm): Ar (50sccm), another is SF_6 (4sccm):C_2H_4 (2sccm): Ar (50sccm). The etching rate is 87 and 100 nm/min which is suitable to the depth to be etched.

TABLE I. Result of silicon etching under different gas flux rate

Gas Flux Rate(sccm)			Silicon Etching Rate(nm/min)	Ratio of Etching Rate(Si/resist)	Sidewall Roughness	Vertical Sidewall or not
SF_6	C_2H_4	Ar				
40	0	0	750	1.6	high	not
20	0	0	600	2.0	high	not
10	0	0	450	2.5	high	not

Ex. 1016, Page 1095

10	1	0	420	5.0	high	not
10	2	0	380	7.2	high	not
4	2	50	87	2.3	low	vertical
3	1	50	100	2.2	low	vertical

Optimization of RF/LF power

Table.2 lists the results of silicon etching under different LF/RF Power. When LF power was fixed to 300W and RF Power was chosen to under 5W, the roughness of the etched sidewalls is very high and the shape of the waveguide cross section is not rectangular. As increasing the RF power to 20W, a waveguide with low roughness and vertical sidewalls can be obtained. This can be explained as follows: RF power provides plasma high energy towards the sample. When RF power is low, the plasma radicals can not get enough energy and the moving direction is scattered, so anisotropic silicon etching can hardly be obtained. As RF power reaches to 20W, the plasma radicals with high energy and vertically shoot to the substrate. These conditions support the anisotropic silicon etching.

We also investigate the influence of the LF power by fixing the RF power to 20W and choosing the LF power to 300W, 350W, 400W. Ultra low roughness was obtained at LF choosing to 300W. Another optimized result with SF_6 (4sccm):C_2H_4 (2sccm): Ar (50sccm) is also investigated, the most satisfied waveguide has been obtained at RF power to 20W and LF to 250W.

TABLE II. Result of silicon etching under different LF/RF Power

Gas Flux Rate(sccm)			**LF**	**RF**	**Silicon Etching**	**Sidewall**	**Vertical**
SF_6	C_2H_4	Ar	**Power(W)**	**Power(W)**	**Rate(nm/min)**	**Roughness**	**Sidewall or not**
3	1	50	300	3	53	high	not
3	1	50	300	5	67	high	not
3	1	50	300	20	105	Ultra low	vertical
3	1	50	350	20	109	low	vertical
3	1	50	400	20	108	low	vertical
4	**2**	**50**	**250**	**20**	**69**	**Ultra low**	**vertical**
4	2	50	250	10	23	Ultra low	vertical

Silicon nanowires and ring resonators

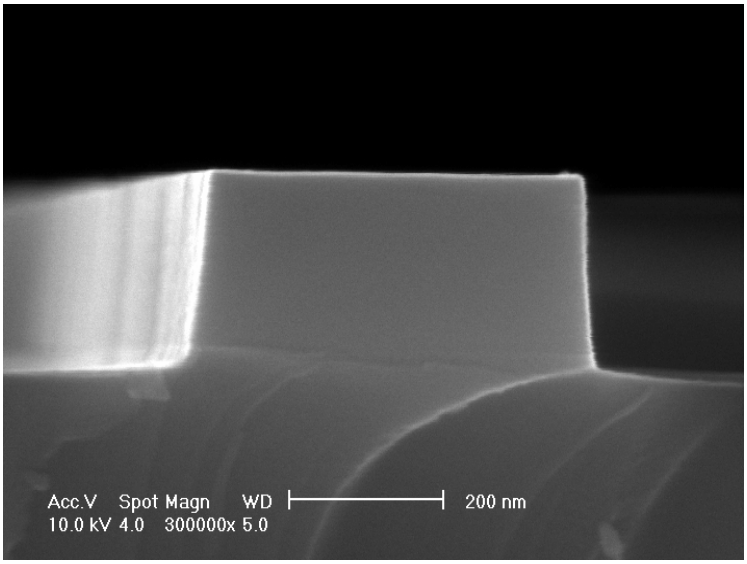
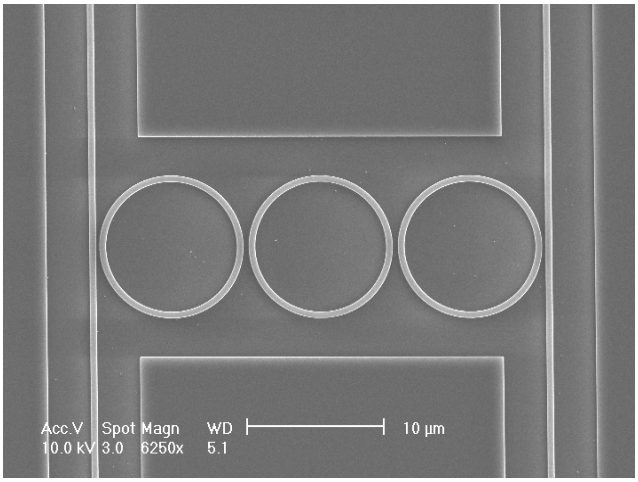
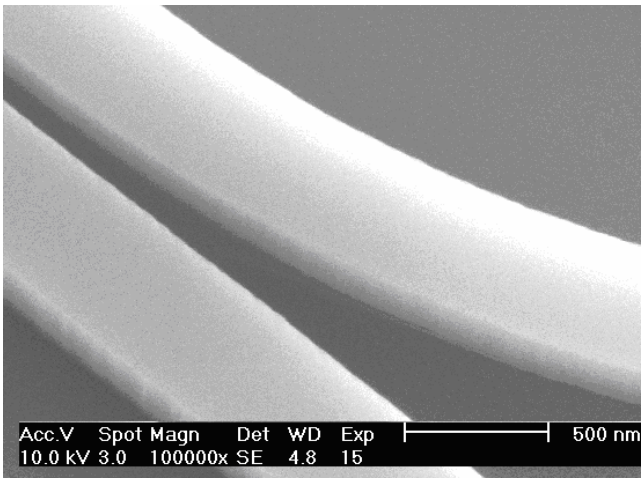


Figure 2. The SEM micrographs of silicon nanowires' cross section

Fig. 2 shows the cross section of silicon nanowire, the etching condition is listed in the table II which is signed in bold. The bottom angle is 86° , nearly vertical sidewalls have been obtained. They are also quite smooth. For absenting measurement equipments, we can not measure the roughness of sidewalls. Fig. 3 shows the resonators which we have fabricated using the above etching condition. Fig. 3(a) shows the top view of the device and Fig. 3(b) shows the gap between waveguide and ring. The size of the gap which we have designed is 130nm, after electron beam lithography and ICP etching, the size of the gap enlarged to 200nm, so we can obtain the comparative precise size by design smaller gap to compensate consume of the ICP etching. The device exhibits typical resonator response characteristic.



(a)



(b)

Figure 3. The SEM micrographs of (a) serially coupled triple ring resonator. (b) The gap between waveguide and ring.

Conclusion

In this paper, a silicon dry etching technology using ICP with $SF_6/C_2H_4/Ar$ was reported. Through the optimization of the major factors such as the flux rates of the etching gases, the source power and bias power, silicon nanowires with vertical and smooth sidewalls are achieved combined electron beam lithography. By designing appropriate offsets, the generally accurate device size can be obtained. Ring resonator with 200nm gap was successfully fabricated, device exhibits typical resonator response characteristic. This technology is suitable for fabrication of planar photonic integrated components and circuits.

Acknowledgments

The authors acknowledge Prof. Ke Wei for useful suggestions.

References

1. Kimmo Solehmainen, Timo Aalto, *J. Lightwave Technology.*, **23**, 3875 (2005).
2. Chen, S.C, Kuo, C.Y, Lin, Y.C, Wu, J.C, Hong, L. in *Emerging Technologies - Nanoelectronics, 2006 IEEE Conference, on Volume , Issue , 10-13 Jan. 2006 Page(s): 467 – 471*.
3. Jingtao Zhou, Huajun Shen, Huihui Zhang, Xinyu Liu, *The 11th IEEE International Conference on Communications Systems/2008.*, p.1231,Optical Systems and Photonic Devices, Guangzhou(2008).

ECS Transactions, 18 (1) 1071-1074 (2009)
10.1149/1.3096574 © The Electrochemical Society

The synthesization of Si nanocrystals embedded in HfO2 matrix by electron beam co-evaporation technique

Weilong Li[a], Rui Jia[a], Chen Chen[a], Ming Liu[a], Haofeng Li[a], Chenxin Zhu[a], Shibing Long[a]

[a] Key Laboratory of Nanofabrication and Novel Devices Integrated Technology, Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China

Silicon nanocrystals synthesized by electron beam co-evaporation (EBCE) of Si and HfO_2 mixture are studied. HfO_2 and Si powder are uniformly mixed together in certain proportion. The mixed Si & HfO_2 powder is evaporated by electron beam evaporation technique. Then the SRO thin films are annealed at different temperatures at 2 hours in N_2 ambient to synthesize Silicon nanocrystals. For the sample annealed at 1050 ℃ at 2 hours, silicon nanocrystals with different size between 3 nm to 5 nm, the mean diameter of 4.0 nm, and the density of 3×10^{12} cm^{-2} are evidently observed by high resolution transmission electron microscopy (HRTEM). Then the Raman scattering and photoluminescence spectra arising from Silicon nanocrystals are further confirmed the above result. In addition, the samples annealed at 900 ℃ for 2 hours to 8 hours are carefully studied by Raman spectra. This way of preparing SRO thin films and Si NCs has unique advantage and is controllable and flexible.

1. Introduction

Recently, Silicon nanocrystals (Si NCs) embedded in high-κ matrix have attracted much attention for their potential applications in nanocrystal nonvolatile memory (NC-NVM) and optoelectronic devices (1-4). They can be produced by annealing Si-rich dioxide or oxide (SRO) thin films in an inert ambient. To date, several techniques have been employed to deposit SRO thin film, including plasma enhanced chemical vapor deposition (PECVD), co-sputtering, and laser ablation of Si in ambient oxygen. Besides these techniques, electron beam evaporation is an attractive and efficient way of evaporating a source material and allowing for thin film coatings upon a substrate. This process is compatible with standard silicon integrated circuit technology, and can obtain hydrogen-free thin film.

2. Experimental

We firstly present a method for the synthesization of Si NCs embedded in HfO_2 by using electron beam co-evaporation (EBCE) to evaporate Si & HfO_2 mixture and to deposit SRO thin films.

In our experiment, firstly, HfO_2 and Si powder are uniformly mixed together in certain proportion. The mixed Si & HfO_2 powder is evaporated by electron beam evaporation technique. The deposited silicon rich dioxide film thickness on the substrates is 300 nm. The film thickness can be precisely controlled by monitoring the evaporation rate and time. After the deposition, samples were annealed in N_2 ambient at different temperature ranging from 700 ℃ to 1050 ℃ for different time. Transmission electron

Ex. 1016, Page 1099

microscopy (TEM) measurement, photoluminescence (PL) and Raman scattering were used to characterize the synthesization of Si NCs embedded in HfO_2 matrix.

3. Results and disscussion

To synthesize Si NCs, we annealed the SRO samples from 700 ℃ to 1050 ℃ for 2 hours. Then the samples were investigated by cross section TEM.

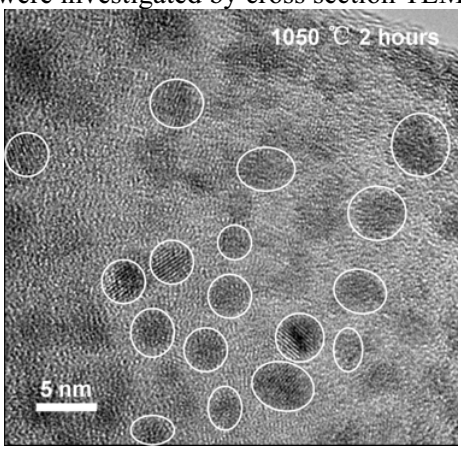


Figure 1. TEM image of Si NCs embedded in HfO_2 matrix.

Figures 1 shows the cross-sectional TEM images of the samples annealed at 1050 ℃, high density Si NCs can be clearly seen. They are uniformly distributed in the HfO_2 matrix and their size ranges between 3 and 5 nm, with a mean diameter of ～4.0 nm and density of 3×10^{12} cm^{-2}, which is high enough for the NVM application.

In order to further confirm the above results, Raman spectra of as-deposited and annealed samples were measured in a backscattering geometry by using the Dilor Super Labram Micro-raman system with wavelength of excitation laser light being 488.0 nm. The results are shown in Fig. 2.

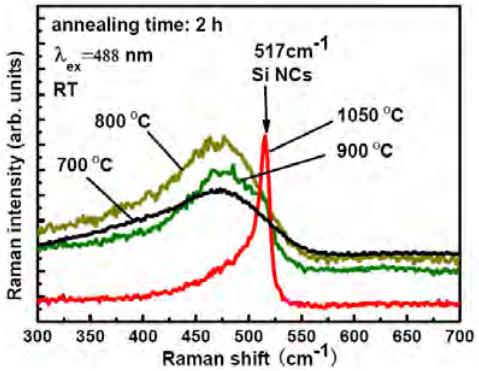


Figure 2. Raman spectrum for SRO films annealing at different temperature for 2 hours.

For the samples annealed at 700 ℃, 800 ℃, and 900℃, no clearly sharp peak is observed with only broad and asymmetric bands centered at about 480 cm^{-1} (TO mode), which are regarded as the typical features of amorphous Si (α-Si) nanoclusters, indicating that phase separation takes place even at 700 ℃.

While increasing the annealing temperature, the FWHM (Full Width at Half Maximum) of the broad peak centered at 480 cm^{-1} becomes smaller, and the Raman peak sharpens with the increasing of temperature, strongly suggesting an increase in the

ordering of the α -Si nanoclusters and transition from amorphous to crystalline nanocrystals. When annealed @ 1050 ℃, a sharp peak appears at about 517 cm^{-1} with a relative weak tail at 480 cm^{-1}. As reported by other people (5), the appearing of sharp peak is the evidence of forming Si NCs, which is also in good agreement with TEM observation as shown in Fig. 1. Here, the Raman shift can be predicted using the Bond Polarisability Model-ling (6) and it can be expressed as the expression: $\Delta\omega=\omega(d) - \omega 0 = -A(\alpha / L)\gamma$, where ω(d) (517 cm^{-1}) is the frequency of the Raman phonon in a nanocrystal with size L̃, ω0 (～520 cm^{-1}) is the frequency of the optical phonon at the zone center for bulk crystal Si, and α is the lattice constant of Si (0.543 nm). The parameters A and γ, used to describe the vibration confinement due to the finite size in a nanocrystal, are 47.41 cm^{-1} and 1.44, respectively. After the calculation, the L is about 3.7 nm, which is consistent with the TEM observation of 4 nm.

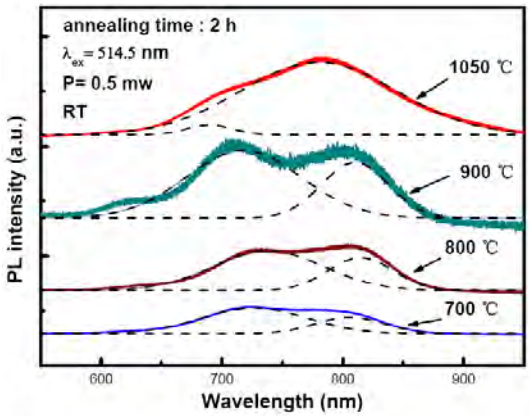


Figure 3. PL spectra for SRO films annealing at different temperature for 2 hours.

Meantime, the PL spectra were measured and shown in Fig. 3. We can see that for the samples annealing at 700 ℃, 800℃, and 900 ℃, there have two broad PL peak located at about 720 nm and at 816 nm, the former of which may be caused by the compound of Hafnium Silicate, but the later of which is believed to be caused by the α -Si. However, for the sample annealing at 1050 ℃, there has one powerful PL peak at about 800 nm, which obviously mainly arises from the silicon nanocrystals; the relative weak peak at about 700 nm may be also caused by the compound of Hafnium Silicate. So, with the increasing of annealing temperature, the amorphous Si (α -Si) are increasing and ordering. And when the temperature up to 1050 ℃, mass silicon nanocrystals are formed. This result is also consistent with the Raman spectra shown in Fig.2.

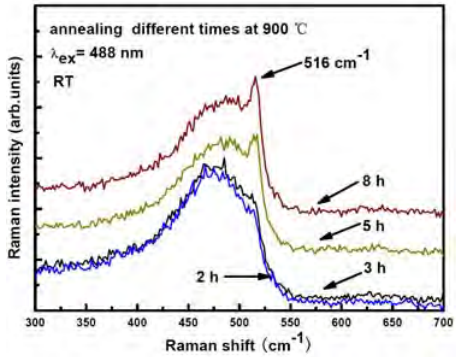


Figure 4. Raman spectra for SRO films annealing at different times at 900 ℃.

In order to clarify the silicon phase transition, the Raman spectra annealing at 900 ℃ from 2 hours to 8 hours were also measured and shown in Fig. 4. It can be seen that with increasing annealing time, the phase transition from α-Si to Si NCs also occurred, since annealing for 5 h, a sharp Raman peak at 516 cm-1 was observed. When prolonged to 8 h, the peak was evident and sharp. However, the broad TO Raman peak centered at 480 cm^{-1} also exists, indicating the co-existence of α-Si and Si NCs.

4. Conclusions

In summary, we firstly use EBCE of the Si and HfO_2 mixture to deposit SRO films. After annealing SRO films at different temperatures and time, their TEM and optical properties were studied. We evidently synthesized Si NCs embedded in HfO_2 matrix. This way of preparing SRO thin films and Si NCs has unique advantage and is controllable and flexible

ACKNOWLEDGMENTS

This work is subsidized by the 973 Projects 2006CB604904, 2006CB302706, by the Chinese NSF under Grant Nos.60706023, 60676001, 90401002, 90607022, by CAS YZ0635 and by CAS 'Bai Ren Ji Hua'

References

1. Sandip Tiwari et al, Appl. Phys. Lett. **68,** 1377(1996).
2. L. Pavesi et al, Optics laser in Engineering **39**, 345 (2003).
3. A M. L. Brongersma et al, Appl. Phys. Lett. **72**, 20 (1998).
4. R. Timothy et al, IEEE Trans. Nucl. Sci. **52**, 6 (2005).
5. C. Meier et al, J.Appl. Phys. **101**, 103112 (2007)
6. Jian. Zi et al, Phys. Rev. B. **58**, 6712 (1998)

ECS Transactions, 18 (1) 1075-1080 (2009)
10.1149/1.3096575 © The Electrochemical Society

New Structure Improves Copper Phthalocyanine (CuPc) Order and Decrease the Device Resistance

Ge Liu, Ming Liu, Liwei Shang, Zhuoyu Ji, Xinghua Liu, Hong Wang and Jiang Liu
Laboratory of Nanofabrication and Novel Device Integration,Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China

We fabricated high performance organic field-effect transistors with novel hybrid electrode-semiconductor contact geometry. The HC OFETs contain both the top electrodes and the bottom electrodes in a single transistor simultaneously. The bottom electrodes induce an ordering CuPc film during the active layer depositing. Due to a high order CuPc layers and low total device resistance, as well as little charge effect below the top electrodes, OFETs with the hybrid contact geometry exhibit a higher performance. With convenience of the new structure, the contact resistance between the two traditional contact geometries was studied in one single transistor.

Introduction

In the past twenties years, great development in organic field-effect transistors (OFET) have a substantial impact on both in the industrial and academic worlds (1). Advancement in organic field-effect transistor technology has typically focused on development of device performance, especially for high charge carrier field-effect mobility (μ_{FE}) (2). Nowadays, dielectric surface modification with self-assembled monolayers (3) or chemical modification of electrodes (4) has been mostly confined to improve the OFETs performance. Recently, by employing a prepatterned substrate to induce a better orderly organic thin film, a new methodology has brought out to optimize the devices (5-7). The technologies for prepatterning the substrate include directionally rubbed and assembled nanoribbons or other designs on the substrate. These works revealed that the factitious ridges on the substrate surface were able to increase the polymer or small molecule ordering and then the transistors' performance (8).

In this paper, by means of depositing gold lines on the substrate before organic molecules deposition, we report a simple fabrication to pattern the substrate. The gold lines were patterned by conventional photolithography and lift-off techniques. The most interesting aspect of this process is about the action of the gold lines. They served as inducing precursor during the course of semiconductor deposition. After the top electrodes were deposited, the gold lines interconnect with the top electrodes spontaneously and then worked as a comb-shape bottom electrodes. As a result, the transistors were fabricated in a hybrid contact (HC) geometry, which contained top electrodes and bottom electrodes in one single transistor simultaneously.

Experimental

The schematic diagrams of hybrid contact organic field-effect transistor are

Ex. 1016, Page 1103

shown in Fig. 1(a), and the corresponding scanning electron microscopy (SEM) image is shown in Fig. 1(b). HC OFETs were fabricated on a highly doped n^+-Si wafer with 300 nm thermally grown oxide layer. Initially, two groups of 30 nm thick gold interleaved lines were patterned and deposited by conventional photolithography. The gold lines are 60 μm long with a width range from 2 um to 8 μm in different transistors, and the minimum gap at the interleaved region was 2 μm. Subsequently, copper phthalocyanine (CuPc) was thermally evaporated with a thickness of 50 nm.

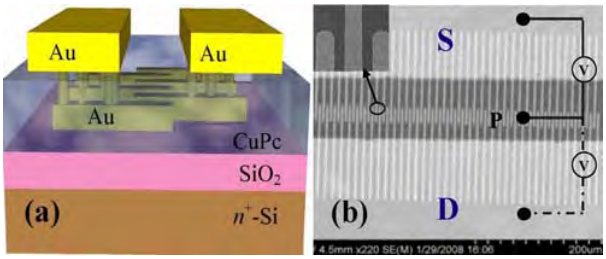


Figure 1. (a) Schematic diagrams of hybrid-contact OFET; (b) the SEM images of a hybrid-contact OFET, the left inset is the zoomed SEM image in the channel and the right inset are schematic probe positions during the I-V test.

Finally, 50nm thick gold source and drain electrodes were formed on the CuPc film by e-beam evaporation through a metal shadow mask, which was manipulated under microscopic observation.

During the top electrodes depositing, the tiny gold particle penetrated through the active layer (9) and formed vast thin conductible filaments. Consequently, the top electrodes and the bottom gold lines were connected spontaneously. So the HC OFETs were fabricated. I-V tests showed that the top electrodes and the relevant bottom electrodes were in short circuit, namely, interconnected. The schematic test points were shown in right inset of Fig. 1(b). For comparison, ordinary top contact and comb-shaped bottom contact OFETs were also fabricated with the same process.

Results and discussion

Electrical measurements were carried out at room temperature in air condition by use of a Keithley 4200-SC system. Fig. 2 shows the typical drain output and transfer characteristics of the HC OFETs. In the Fig. 2(a), we can observe a clear linear drain current regime and a saturation regime with the source-drain voltage increasing at different gate voltage step. It implies the transistors hold a good Ohmic contact between the gold electrodes and the organic semiconductor (10). The output characteristics fit the standard source-drain current versus voltage equations well (11). From the square root of the drain current versus gate voltage curve at a fixed source-drain voltage of -40V, the field-effect mobility is estimated to be 0.02−0.05 $cm^2V^{-1}s^{-1}$. The threshold voltage (V_T) and on/off ratio determined from Fig. 2(b) are about -6.3V and 10^4. Compared with the report by Zhenan Bao (12), the HC OFETs nearly double the mobility of Cupc thin film and decrease the threshold voltage at an extent of -4V.

In order to understand the role of the new device structure, traditional top contact

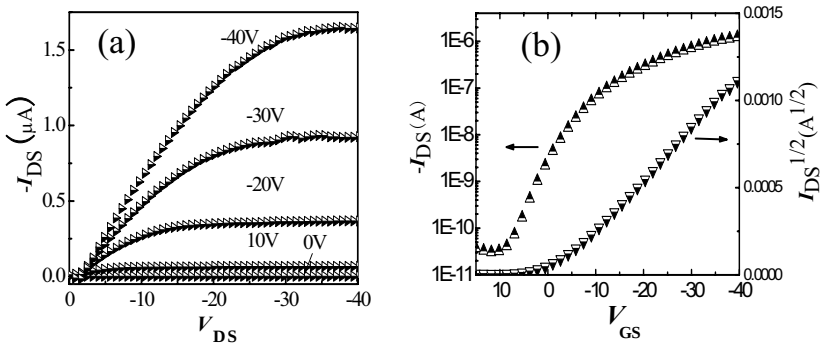


Figure 2. Show drain output (a) and transfer (b) characteristics of the HC OFET, respectively.

and bottom contact OFETs were also fabricated in the same process. Fig. 3 shows the output characteristics of the OFETs with common top contact [Fig. 3(a)] or bottom contact geometry [Fig. 3(b)]. The even mobility and threshold voltage are about 0.018 cm^2/Vs and -8V for the TC geometry while 0.00093 cm^2/Vs and -13V for the BC one. The on/off ratios are 10^4 and 10^3, respectively. Compared with the traditional TC and BC geometries, the HC geometry improves the mobility and decreases the threshold voltage.

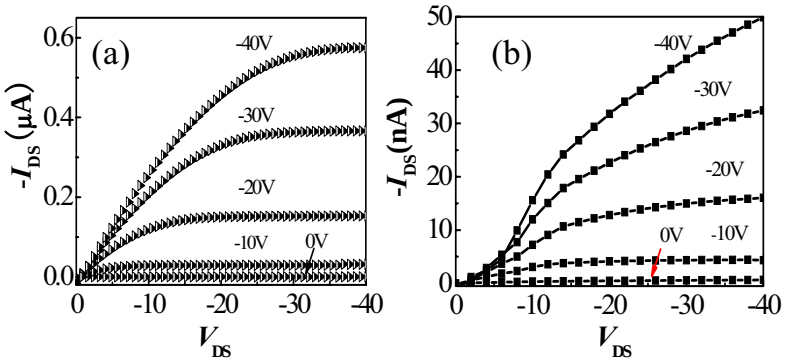


Figure 3. Output characteristics of the conventional TC OFETs (a) and BC OFETs (b).

It is well known that an ordering organic thin film means a less defect and then a higher carrier mobility. Besides the depositing parameters, a patterned substrate is also an optional means to improve the organic molecule order (5-8). Takes the literature by Masato Ofuji *et al.* (13) for example, by use of rubbing treatment substrate, the CuPc layers were induced in uniaxial orientation, and then got a higher hole mobility of CuPc. It suggests the ridging substrate is able to improve the ordering

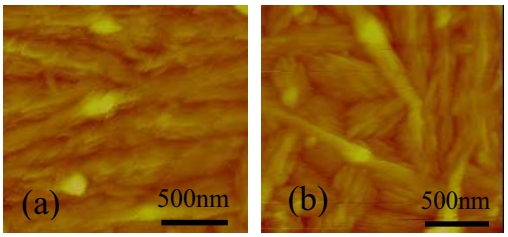


Figure 4. AFM images of the CuPc films' morphology on SiO_2 substrate with the bottom induceing lines (a) and without (b).

of CuPc film. In the HC OFETs, the bottom gold lines could take a similar effect on the CuPc growing as the rubbing scrapes. Fig. 4 shows the AFM image of CuPc with bottom gold lines [Fig. 4(a)] or without [Fig. 4(b)]. The bottom lines help to improve the CuPc films orientation, then decrease the grain boundaries scattering for the hole carriers. This would be taken into account for the higher mobility and lower threshold of the HC OFETs.

Simultaneously, within the HC OFETs, the top electrodes and the bottom electrodes are combined in parallel circuit between the sources and drain electrodes, which lead to a lower total equivalent device resistance than the single top contact or bottom contact OFETs. Additionally, the hybrid electrodes would change the electrical field distribution in the channel, and especially, suppress the charging effect (14) below the top electrodes because the top electrodes and the bottom electrodes nearly keep the same potential. Its benefit the carriers' injection at the electrode-organic interface and transportation in the channel.

For the BC OFETs, as shown in Fig. 3(b), the super-linear and unsaturated behavior occurred with the drain voltage increasing imply a high contact resistance (15) and an evident short-channel effect (16). Which are also considered to limit the drain-current flowing through the bottom electrodes in the HC OFETs. As a result, the top electrodes play a dominant role versus the bottom electrodes in the HC OFETs, as shown in the Fig. 5.

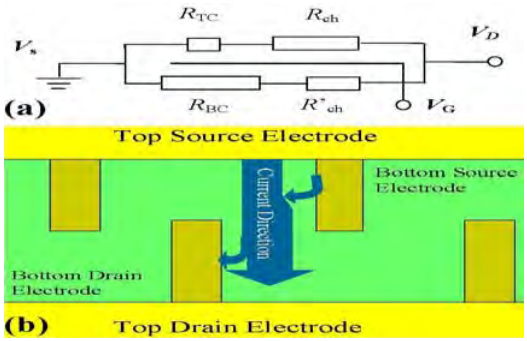


Figure 5. (*a*) The top electrodes and bottom electrodes were combined in parallel resistances. R_{TC}, R_{ch},were the contact resistance and channel resistance for the top electrode, respectively; R_{BC}, R'_{ch} were the contact resistance and channel resistance for the bottom electrode, respectively; (*b*) schematic of the current direction in hybrid OFETs.

The bottom and top electrodes are fabricated with the same process and operated under the same electrical parameters, while the contact resistance or total resistance is remarkably different from each other. It is well known that the charge carrier injection intensively depend on the energetic barrier and the effective injection area. In the case of HC OFETs, the top and bottom electrodes are fabricated in the same process simultaneous and keep the same potential at any time. Therefore the factors which relate to the work function difference or result in a variety of the Fermi energy are able to ignore.

By studied the interface morphology between the electrodes and the organic material, we consider that the different growing style between the metal electrodes

Ex. 1016, Page 1106

depositing on the organic semiconductors and the organic materials growing on the prepatterned bottom electrodes could be responsible for the obviously different contact resistance. The influences are illuminated in the following aspects.

Firstly, it results in a bigger contact area for the top electrodes than bottom electrodes. Big effective contact area implied a small contact resistance (17). As shown in Fig. 6(a), because of the extraordinary roughness of the organic-electrode interface, the real contact area between the top electrodes and the organic layer was bigger than the planar size of the top electrodes. While for the BC geometry, the organic materials couldn't completely cover the surface of the electrodes [(see in Fig. 6(b)]. This implied the effective injection area between the CuPc and bottom electrodes is much smaller than that of the TC configuration devices, which result in a bigger contact resistance in the BC configuration devices. Secondly, the anisotropic mobility of the organic semiconductors would also add an additional resistance to the bottom electrodes. Prior to the organic layer deposition, bottom electrodes have been

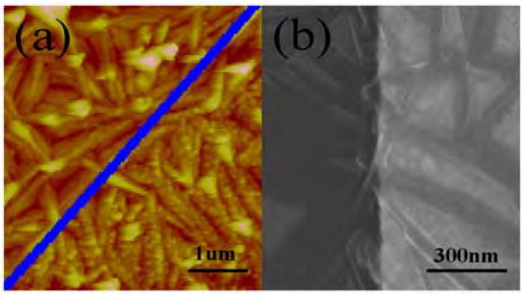


Figure 6. (*a*) AFM image of gold top electrode (right) on CuPc (left); (*b*) SEM image of CuPc on the edge of gold bottom electrode (right).

fabricated on the dielectric surface, and their sides are obviously vertical to the substrate. Therefore, for the bottom electrodes, the initial several molecule layers on the electrode sides and the nearby dielectric surface are out of order [see in Fig. 6(b)]. Some literatures had found that the organic molecule would stand on the surface of poly gold and silicon dioxide in the course of deposition (18). That is equivalent to insert a thin film between the electrodes and the active layer, and subsequently increase the resistance of the bottom electrodes.

Conclusions

We fabricated high performance organic field-effect transistors with novel hybrid electrode-semiconductor contact geometry. The HC OFETs contain both the top electrodes and the bottom electrodes in the single transistor. The bottom electrodes not only serve as bottom electrodes, but also help to induce an ordering CuPc layers during the active depositing. Due to a high order CuPc layers and low total device resistance, as well as little charge effect below the top electrodes, the hybrid contact geometry improves the OFETs performance. In addition, we study the contact resistance and consider the contact area and anisotropic mobility are taken into account for the high contact resistance in bottom contact geometry.

Ex. 1016, Page 1107

Acknowledgments

This work was supported by National Basic Research Program of China (973 Program) under Grant No: 2006CB806204 and Natural Science Foundation of China under Grant No: 60676001, 60676008 and 60825403.

Reference

1. S. R. Forrest, *Nature (London)* **428**, 911 (2004).
2. H. Sirringhaus, *Adv. Mater. (Weinheim, Ger.)* **17**, 2411 (2005).
3. A. Facchetti, M.-H. Yoon, and T. J. Marks, *Adv. Mater. (Weinheim, Ger.)* **17**, 1705 (2005).
4. Chong-an Di, Gui Yu, Yunqi Liu, Xinjun Xu, Dacheng Wei, Yabin Song, Yanming Sun, Ying Wang, Daoben Zhu, Jian Liu, Xinyu Liu, and Dexin Wu, *J. Am. Chem. Soc.* **128**, 16418 (2006).
5. Xiangfeng Duan, Chunming Niu, Vijendra Sahi, Jian Chen,J. Wallace Parce, Stephen Empedocles & Jay L. Goldman, *Nature (London)* **425**, 274 (2003).
6. Gregory L. Whiting, Henry J. Snaith, Saghar Khodabakhsh, Jens W. Andreasen, Dag W. Breiby, Martin M. Nielsen, Neil C. Greenham, Richard H. Friend, and Wilhelm T. S. Huck, *Nano lett* **6(3)**, 573 (2006)
7. G. Derue, D.A. Serban, Ph. Leclère, S. Melinte, P. Damman, R. Lazzaroni, *Org. Electr*. **9**, 821 (2008).
8. Patrick B. Shea, Charlene Chen, and Jerzy Kanicki, *Appl. Phys. Lett.* **90**, 233107 (2007).
9. M. Abkowitz, J. S. Facci, and J. Rehm, *J. Appl. Phys*. **83**, 2670 (1998).
10. T. J. Richards and H. Sirringhaus, *J. Appl. Phys*. **102**, 094510 (2007).
11. Zhenan Bao, Andrew J. Lovinger, and Ananth Dodabalapur, *Appl. Phys. Lett*. **69**, 3066 (1996).
12. Gilles Horowitz, *Adv. Mater. (Weinheim, Ger.)* **10**, 365 (1998).
13. Masato Ofuji, Ken Ishikawa, and Hideo Takezoe, *Appl. Phys. Lett*. **86**, 062114 (2005).
14. Y. Roichman and N. Tessler, *Appl. Phys. Lett*. **80**, 151 (2002).
15. Chong-an Di, Gui Yu, Yunqi Liu,and Daoben Zhu, *J. Phys. Chem. B*. **111**, 14083-14096 (2007).
16. Joshua N. Haddock, Xiaohong Zhang, Shijun Zheng, Qing Zhang, Seth R. Marder and Bernard Kippelen, *Org. Electr*. **7**, 45 (2006).
17. R. A. Street and A. Salleo, *Appl. Phys. Lett.* **81** 2887 (2002).
18. Wei Chen, Shi Chen, Han Huang, Dong Chen Qi, Xing Yu. Gao, and Andrew Thye Shen Wee, *Appl. Phys. Lett.* **92** 063308 (2008).

ECS Transactions, 18 (1) 1081-1086 (2009)
10.1149/1.3096576 © The Electrochemical Society

Investigating the Effects of Different Gate Biasing Voltages and Temperatures on RadFETs

Hao Tang[a b], Yichen Liu[a b], Yi Wang[a], Jinyan Wang[a], and Yufeng Jin[a b]

[a] Institute of Microelectronics, Peking University, Beijing 100871, P. R. China
[b] Peking University Shenzhen Graduate School, Shenzhen 518055, P. R. China

This paper mainly focuses on the effects of different gate biasing voltages and temperatures on several kinds of RadFETs. Four kinds of gate biasing configurations (positive, zero, negative gate voltage and continuous current applied model) have been compared in terms of sensitivity to radiation. Post-irradiation characteristics of annealing or fading under different gate biasing voltages and temperatures are also studied. Samples irradiated by zero and negative voltage biasing have lower sensitivity than that of positive, but don't show significant fading. It is presented that higher temperature can promote the process of annealing.

Introduction

Some researches have indicated radiation energy can bring MOS devices positive charges located in gate oxide layer (1) (2). As the absorbed radiation dose increasing, these positive charges are gradually accumulated (3) (4). When the number of charges reaches at a certain extent, MOS devices will be damaged. PMOS Radiation sensitive Field Effected Transistor (RadFET) has been widely used as dosimeter in spacecraft, medicine and nuclear industry (5). By measuring the drift of threshold voltage ($\triangle V_{TH}$) of RadFET under radiation condition, we can figure out the amount of radiation dose absorbed. A readout circuit configuration for measuring V_{TH} is showed in Figure 1, I_{Curr}=10uA (6).

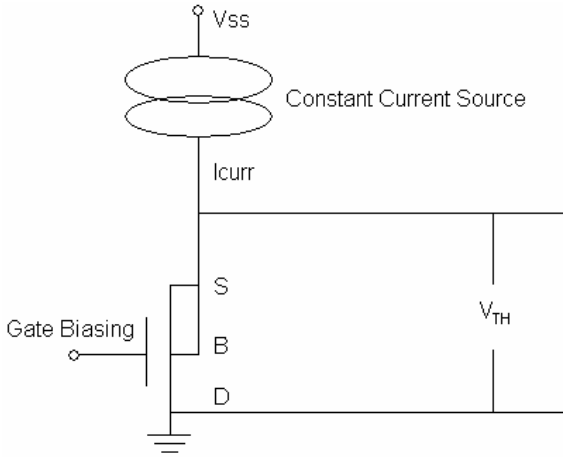


Figure 1. Reader circuit configuration for threshold voltage measurements in RadFET

It is a consensus that sensitivity and stability are two critical parameters for RadFET, but how to increase sensitivity while maintaining stability still needs more improvements up to now (7). This paper reported some investigations focusing mainly on Gate biasing voltage and temperature. Both of them have great influences on the drift and fading of

$\triangle V_{TH}$ which is directly related to the sensitivity and stability (8). Several RadFETs with Wet/Dry/Wet oxidation layer were fabricated. Figure 2 is a TEM picture used to demonstrate the thickness of the gate oxide is 310nm.

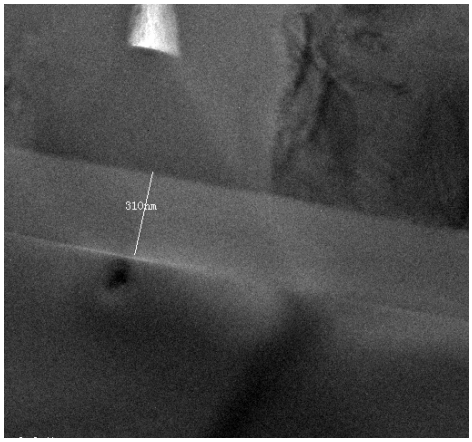


Figure 2. A TEM picture of the gate oxide of RadFET

A particular automated read-out system was also designed for real-time radiation dose measurement. Characteristics of these RadFETs applied with different gate biasing voltages and temperatures have been compared in terms of radiation sensitivity, long-term fading and dose rate dependence. We found that positive gate biasing (+5V) has better sensitivity than zero and negative biasing (-5V), however more obvious fading which reduces stability was also detected. In this paper, the fading of $\triangle V_{TH}$ is also analyzed as a function of ambient temperature.

Experiments

These PMOS RadFETs are processed in Key Laboratory of Microelectronic Devices and Circuits, Institute of Microelectronics, Peking University. All the samples are fabricated on orientation <100>, resistivity 2～4Ω·cm, n type silicon. Gate oxides of RadFET are grown by Wet/Dry/Wet oxidation processes and annealed in nitrogen ambient under two kinds of temperature, 950℃ and 1050℃. Both implanted and unimplanted RadFETs are prepared for the experiment. Boron implantation is applied through oxide to adjust the threshold voltage approaching to zero. The implantation energy 130keV, which is chosen for convenient calculation for initial V_{TH}, is the exact energy for peak value of boron ion preserved at the interface between silicon and gate oxide.

These RadFETs were irradiated up to total dose 30krad by the Co-60 source at a dose rate of 5rad/s. $FeSO_4$ dosimeters are used for dose measurement. After irradiation, fading phenomenon of RadFETs is observed in constant temperature chest in a temperature range of -15℃ to 40℃. The read-out system shown in Figure 3 is the most important instrument in the radiation experiment. RadFETs are exposed to radiation source, and the reader circuit is shielded by lead brick. Before this read-out system was designed, RadFETs would be irradiated, taken out of irradiation chamber after irradiation, and then measured. To solve this problem, the four testing models can be easily changed in on-line measurements, and real-time reading output signals are realized for experiment

Ex. 1016, Page 1110

requirements. Three voltage biasing models are not applied on constant current source during irradiation, and constant current model is used for measuring output voltage.

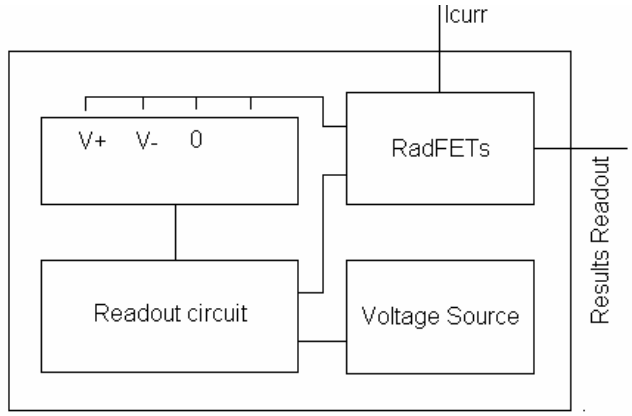


Figure 3. Automated read-out system designed for real-time measurement

Results and Discussion

The W/L rate of RadFETs is 350/50 (W and L are both in um). Transfer characteristics of unimplanted RadFET using base 10 logarithmic coordinate are illustrated in Figure 4, the initial threshold voltage is about -4.5V, drain voltage V_D are -0.2V and -0.05V.

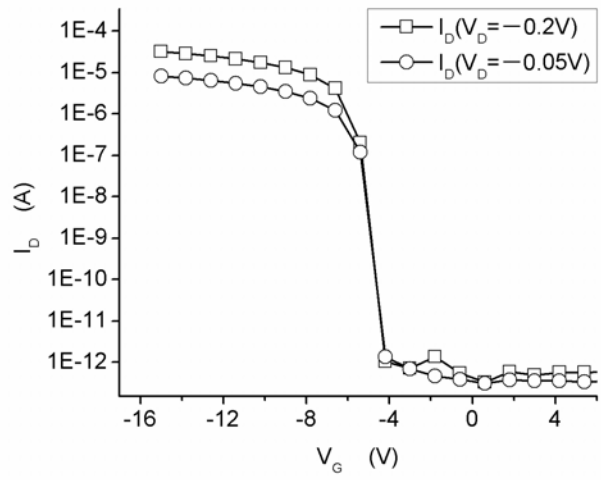


Figure 4. Transfer characteristics of unimplanted RadFET before radiation experiment

Response to radiation curves ($\triangle V_{TH}$ versus dose) can be standardized to a equation

$$\triangle V_{TH} = A * D^B \quad [1]$$

Where D is the total dose in krad, A and B are two parameters obtained by data fitting. The average sensitivity figures in three sections of total dose range are given in TABLE I. Analyzing from equation [1] and TABLE I, it can be concluded that $\triangle V_{TH}$ decreases with dose. That is to say, $\triangle V_{TH}$ would be saturated when D goes to a large enough extent. As expected, gate biasing models have great effect on the sensitivity of RadFETs (9). The $\triangle V_{TH}$ variations of implanted RadFETs are represented in Figure 5. The highest

Ex. 1016, Page 1111

sensitivity is obtained from +5V gate biasing model, and following order is 0V gate biasing, constant current applied model marked as CCAM and -5V gate biasing.

TABLE I. Radiation sensitivity figures for three total dose sections (mV/rad)

Total Dose Section	+5V Biasing	0V Biasing	-5V Biasing	CCAM
0 to 10 krad	0.190	0.128	0.117	0.087
0 to 20 krad	0.124	0.084	0.075	0.056
0 to 30 krad	0.096	0.065	0.057	0.043

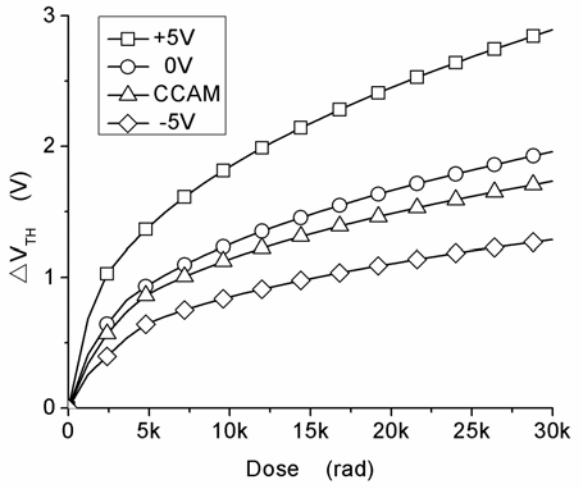


Figure 5. The $\triangle V_{TH}$ variations of implanted RadFETs during radiation

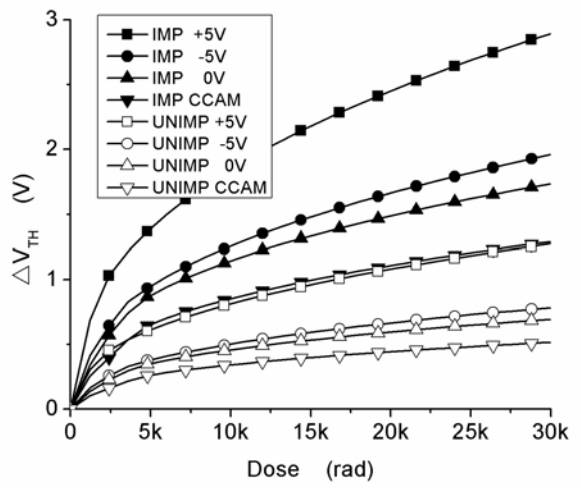


Figure 6. Sensitivities are compared between implanted and unimplanted RadFETs

The variation values of 0V and CCAM are very close, because their configurations are similar. Reviewing Figure 1, Gate is always connected with GND for CCAM. In 0V biasing model, 0V is added to Gate during irradiation, and model is changed to CCAM for reading output voltages. The differences between them are mainly due to the more recovery of holes in CCAM. Positive voltage creating a higher electric field is more helpful for charging holes near the Si/SiO_2 interface. An interesting phenomenon is illustrated in Figure 6. Implanted RadFETs have larger sensitivity than unimplanted RadFETs under either testing configuration. For this issue, one reason is implantation

offers a better scratch line for drift by dragging initial V_{TH} close to 0V, and it also delays the saturation state to enlarge range of RadFETs. The other one is implantation creates additional defects in the gate oxide to enhance the ability of trapping holes (10).

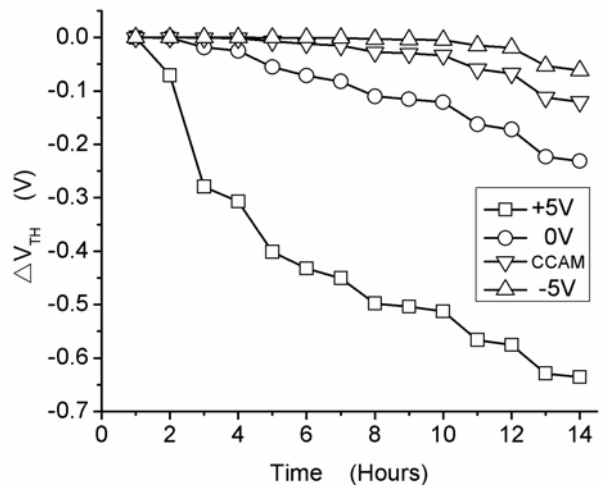


Figure 7. Fading curves of implanted RadFETs

Fading curves used to weigh RadFETs stability are represented in Figure 7. +5V biasing which obtains the highest sensitivity also has the largest fading value (11). This work is proved that fading variations of implanted RadFETs are more obvious than that of unimplanted ones. Actually, fading is a course of electron hole recombination. More holes are collected in gate oxide layer to increase the sensitivity, larger opportunity of recombination occurs (12). Temperature is a significant parameter of fading, because it helps the recombination by promoting the thermal motion of electrons and holes. Figure 8 shows $\triangle V_{TH}$ is shifting with temperature.

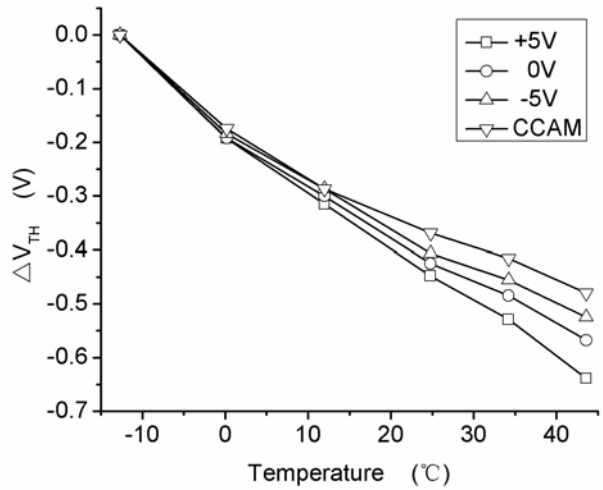


Figure 8. $\triangle V_{TH}$ variation versus temperature

Conclusions

This work included following aspects: implanted and unimplanted RadFETs with

Ex. 1016, Page 1113

300nm thick gate oxides are fabricated, a real-time readout system is designed and manufactured, experimental data are processed and discussed after radiation. The readout system which can change biasing configuration conveniently brings more measurements to radiation experiment. From average sensitivities calculated for three ranges of total dose in Table I, it is proved that Slope of $\triangle V_{TH}$ reduced with dose increase. Equation [1] can explain the saturation of RadFET. Positive voltage creating a higher electric field is more helpful for charging holes near the Si/SiO_2 interface, so +5V biasing obtains the highest sensitivity. Difference between zero biasing and CCAM is the time of connecting with readout circuit. Implanted RadFETs have larger sensitivity than unimplanted RadFETs under either testing configuration. Better scratch line of initial V_{TH} and additional defects created by implantation are assumed to be main mechanisms. It is also concluded from fading curves that the testing model which has a higher sensitivity would have a larger fading value. RadFETs operate under a suitable constant temperature could be considered for the stability in further work.

Acknowledgments

The authors would like to thank China Institute of Atomic Energy who offered valuable advice and great help for radiation experiment.

References

1. G. Ristic, A. Jaksic and M. Pejovic, *Sensors and Actuators*, A 63, p. 129 (1990).
2. A. G. Holmes-siedle, L. Adams and G. Ensell, *Radiation and its effects on devices and systems 1st European conference*, IEEE Trans. Nuc. Sci. Proc. p.65 (1992)
3. D. M. Fleetwood, *IEEE Trans. Nuc. Sci.*, Vol. 39, p.269 (1992)
4. N. Stojadinovic, M. Pejovic, S. Golubovic, G. Ristic, V. Davidovic and S. Dimitrijev, *Electronics Letters*, Vol. 31, p. 497 (1995)
5. G. Ristic, S. Golubovic and M. Pejovic, *Sensors and Actuators*, A 51, p. 153 (1996)
6. Y. Kimoto and A. Jaksic, in *24th Internatioal Conference on Microelectronics/2004*, IEEE Trans. Nuc. Sci. Proc., Vol. 2, p. 657 (2004).
7. A. Kelleher, M. O'Sullivan, B. O'Neill and W. Lane, *IEEE Trans. Nuc. Sci.*, Vol. 39, p.342 (1992)
8. A. Jaksic, G. Ristic, M. Pejovic, A. Mohammadzadeh and W. Lane, in *23rd International Conference on Microelectronics/2002*, IEEE Trans. Nuc. Sci. Proc., Vol. 2, p.727 (2002)
9. C. Conneely, B. O'Connell, P. Hurley, W. Lane and L. Adams, *IEEE Trans. Nuc. Sci.*, Vol. 45, No. 3, p.1475 (1998)
10. Tsang cheung, Martin J Buston and Peter K N Yu, *Physics in Medicine and Biology*, 49, p.191 (2004)
11. A. Haran, A. Jaksic, N. Refaeli, A. Eliyahu, D. David and J. Barak, *IEEE Trans. Nuc. Sci.*, Vol. 51, p.2917 (2004)
12. D. J. Moreno, R. C. Hughes, M. W.Jenkins and C. R. Drumm, in *International Conference on Solid-State Sensors and Actuators/1997*, IEEE Trans. Nuc. Sci. Transducers'97, p.1283 (1997)

ECS Transactions, 18 (1) 1087-1094 (2009)
© The Electrochemical Society

Author Index

Ahmet, P.	39, 71
Ai, Y.	99
Akhmetov, V.	995
Al-Jassim, M.	1049
Allan, Z.	15
Allgair, J.	151
Aochi, H.	89
Aoulaiche, M.	651
April, E.	419
Arai, H.	71
Arazi, D.	151
Babu, S. V.	473
Bai, D.	757
Bai, J.	887
Bai, Y.	1019
Baik, K.	299
Bain, M.	9
Baine, P.	9
Bao, J.	315
Bao, T.	169
Barclay, G.	397
Bartholomew, L.	567, 617
Bavin, A.	567, 617
Bei, E.	605
Beica, R.	689
Beyne, E. O.	695
Beynet, J.	275
Bian, A.	601
Bian, Y.	677
Bien, D.	9
Bishop, M.	151
Blackshear, E. D.	701
Borland, J. O.	1059
Brandt, D.	391
Brian, S.	419
Budhraja, V.	1049
Bunday, B.	151
Burggraf, J.	757
Burgstaller, D.	757
Cai, B.	701
Cai, C.	907
Cai, L.	1019
Cai, X.	491
Call, N.	1049
Cao, Y.	511
Carlson, D.	1049
Carroll, M.	887
Cecil, T.	299
Chan, L.	581
Chang, C.	863
Chang, E.	863
Chang, E.	183
Chang, J.	43
Chang, L.	177
Chang, S.	641, 645, 659
Chattopadhyay, K.	269
Chen, A.	601
Chen, C.	93, 1071
Chen, F.	205
Chen, F.	321, 341
Chen, G.	1019
Chen, I. C.	105
Chen, J.	581
Chen, J.	55
Chen, J.	49
Chen, L.	587
Chen, X.	925, 1031
Chen, Y.	523
Chen, Y.	581
Chen, Z.	243
Cheng, S.	881
Cheng, W.	237
Chi, M.	281, 281
Chi, M.	83, 177
Chiu, B.	553
Chiu, C.	511
Chiu, G.	517
Choi, C.	177
Chong, S.	625, 669
Chou, C.	523, 535
Chou, C.	523, 535
Christiaens, W.	707
Chylak, B.	777, 803
Claeys, C.	3, 61
Clauberg, H.	777
Conner, G.	503
Cook, K.	713
Cordes, A.	151

Cott, D. J. 845
Crupi, F. 61
Cui, D. 593, 605

Dai, B. 989
Dai, S. 15, 27
Dai, W. 67
Dam, T. 299
Dawson, D. 169
De Coster, J. 199
De Gendt, S. 845
De Jaeger, B. 61
de Potter, A. 935
De Roest, D. 275
De Stefano, F. 61
De Wolf, I. 199
Dean, K. 575
Delande, T. 453
Deng, D. 27, 117
Deng, H. 77
Devulapalli, B. 989, 1023
Diao, L. 663
Dixit, G. 269
Du, G. 77
Du, L. 77
Du, T. 485
Dupret, F. 935

Eliyahu, A. 453
Eneman, G. 3, 61
Erk, H. 969
Evans, X. L. 9
Evertsen, R. 837

Fang, Z. 287
Farrar, N. 391
Feng, S. 429
Ferns, J. 315
Fiorenza, J. G. 887
Fu, L. 645
Fu, X. 249
Fu, Y. 659
Fujisawa, H. 39
Fukui, K. 733
Fukushima, T. 33
Furukawa, H. 231

Gambino, J. 205
Gamble, H. 9
Gao, B. 447
Gao, R. 605
Gleason, B. 299
Godwin, M. 151
Goh, W. 869
Gonzalez, M. B. 3
Gös, W. 213
Grasser, T. 213
Groeseneken, G. 651
Gschwendtberger, G. 225
Gu, X. 465
Gui, K. 611
Gunjalb, P. 989
Gusakov, V. E. 1013

Ha, J. 763
Hamamoto, N. 1059
Han, B. 645, 659
Han, R. 77, 219
Han, Y. 907
Hao, Q. 1043
Hao, Z. 99
Hashimoto, S. 787
Haspeslagh, L. 199
Hattori, T. 39, 71
Hawkins, J. 517
He, B. 67
He, W. 321, 341
He, Z. 249
Heinzl, R. 213
Hell, E. 713
Hendriks, R. 837
Hennemayer, M. 713
Heylen, N. 453
Hikavyy, A. 3
Hiramoto, T. 55
Hisamatsu, S. 925, 1037
Ho, B. 409
Ho, H. 15, 27, 117
Hong, J. 333
Honh, J. 763
Hsu, A. 553
Hsu, H. 863
Hsu, M. 187
Hsu, Y. 123
Hu, B. 479
Hu, P. 299
Hu, P. 611
Hu, S. 523, 535

Hu, X. 559
Hu, Y. 249
Huang, R. 281
Huang, R. 21, 99
Huang, S. 441
Huang, T. 553
Huang, W. 535
Huang, Y. 237
Hung, A. 15, 27
Hunsche, S. 357
Huwel, W. 707
Hwang, T. 529
Hydrick, J. M. 887

Iwai, H. 39, 71

Janjam, S. 473
Jeong, Y. 55
Jeserer, G. 255
Ji, L. 321, 341
Ji, L. 581
Ji, Z. 895, 1075
Jia, R. 93, 1071
Jiang, J. 677
Jiang, T. 813
Jianhua, J. 15
Jin, Y. 1081
Jing, J. 491, 511, 553
Johnston, S. 1049
Ju, J. 27, 105, 117

Kaczer, B. 651
Kajihara, S. 231
Kakimoto, K. 925, 983, 1031, 1037
Kakushima, K. 39, 71
Kamimura, H. 71
Kan, H. 321, 341
Kanamaru, T. 733
Kang, J. 219, 875
Kangawa, Y. 925, 1037
Kapur, S. S. 959
Katogi, S. 683
Kellens, K. 275, 453
Kenmochi, T. 739
Kettner, P. 721
Kim, H. 479
Kim, K. 151
Kim, T. 763
Kim, T. 763
Kissinger, G. 995
Knight, S. 161
Koga, Y. 1059
Kohler, C. 357
Kojima, T. 33
Kot, D. 995
Koyanagi, M. 33
Kulkarni, M. 989, 1023
Kuo, C. 863
Kweon, Y. 763
Kwon, O. 803

Lam, E. 351
Lathrop, R. 793
Lazerand, T. 567, 617
Leal, J. 769
Lee, B. 857
Lee, D. 625, 669
Lee, K. 803
Lee, S. 605
Lee, S. F. 625, 669
Lee, T. C. 205
Leunissen, L. 453
Li, H. F. 93, 1071
Li, J. 887
Li, L. 447
Li, W. 93
Li, W. 1071
Li, W. 291
Li, X. 581, 605
Li, X. 677
Li, X. 249
Li, Y. 1019
Li, Y. 141
Li, Y. 453
Li, Z. 983
Li, Z. 587
Li, Z. 43
Liang, P. 547
Liao, C. C. 219
Liao, C. 111
Liew, S. 669
Lim, L. 67
Lim, T. 831
Lin, B. 391
Lin, C. 187, 291
Lin, J. 485
Lin, J. 511

Lin, L. 1001
Lin, P. 587, 601, 611, 677
Lin, P. 43
Lin, Y. 587
Lin, Y. 281
Ling, J. 777
Lipscomb, P. 151
Liu, B. 635
Liu, B. 105, 659
Liu, B. 429
Liu, C. 1043
Liu, E. 15, 27, 117
Liu, G. 895, 1075
Liu, H. 21
Liu, J. 895, 1075
Liu, J. 15, 27, 117
Liu, L. J. 925
Liu, L. 875
Liu, L. 983, 1031
Liu, L. J. 625
Liu, M. 49, 93, 141, 895, 901, 1071, 1075
Liu, Q. 49, 141, 901
Liu, S. 141
Liu, W. 559
Liu, X. 219
Liu, X. 875
Liu, X. 447
Liu, X. 983
Liu, X. 67
Liu, X. 895, 1075
Liu, X. 1065
Liu, Y. L. 435, 547
Liu, Y. 907
Liu, Y. 129
Liu, Y. 1081
Liu, Y. 447
Livelo, E. I. 821
Lochtefeld, A. 887
Loix, F. 935
Long, S. 49, 141, 901, 1071
Loo, R. 3
Lu, Y. 291
Lv, H. B. 281
Lypka, C. 757

Ma, X. 1001
Ma, X. 1019
Maehara, N. 1059
Mahulikar, D. 479, 541
Major, C. 887
Markevich, V. 1013
Martinez de Pinillos, J. V. 161
Maruyama, D. 409
Matsumoto, T. 1059
Matsuo, H. 925, 1037
Matthias, T. 721
Medvedeva, I. F. 1013
Meessen, J. 357
Mehta, V. 1049
Mitchell, S. 9
Miyamoto, Y. 863
Miyase, K. 231
Miyazawa, H. 925
Mohammad, K. A. 625
Mollica, R. 503
Montgomery, J. 9
Mountinho, H. 1049
Mountsier, T. 269
Mueller, T. 995
Murin, L. I. 1013

Nagai, A. 683
Nagai, N. 1059
Nagayama, T. 1059
Naik, M. 593, 605
Nakano, S. 925, 1031
Naman, A. 517
Nemoto, T. 465
Neubert, B. 713
Ng, F. 625
Ng, G. 869
Niermeyer, K. 503
Nikolsky, P. 357
Ning, J. 15, 27, 105, 117
Nishijima, M. 33
Nitayama, A. 89
Niu, X. 435, 447, 547
Nooitgedagt, T. 357
Noriha, H. 33
Nozu, M. 787

Obeng, Y. 161
O'Connor, R. 651
Ohmi, T. 465
O'Loughlin, J. 269
Ong, P. 453
Oppert, T. 751

Ouyang, D. 611

Padhi, D. 581
Pai, Y. 523, 535
Pan, J. 219
Pang, L. 299
Pantisano, L. 651
Pargfrieder, S. 721
Park, A. 183
Park, K. 269
Park, S. 763
Patel, V. 701
Peaker, A. R. 1013
Pei, Y. 33
Peng, C. 611
Peng, J. 635
Peng, L. 635
Peng, S. 881
Philipossian, A. 465
Pradhan, A. 269
Puligadda, R. 757

Qian, G. 193
Qiao, Y. 219
Qiu, Z. 373, 381
Que, D. 1001

Rainey, P. 9
Ramachandran, P. A. 915
Ran, R. 405
Rao, V. 567, 617
Regnier, V. 935
Ren, C. 117
Riva Sanseverino, S. 845
Robinson, M. 769
Rochat, N. 275
Rojas, R. V. 821
Rolinsky, R. 935
Romo Negreira, A. 845
Roy, D. 473
Rupnowski, P. 1049
Ruzyllo, J. 263

Sakaida, Y. 409
Sakamoto, R. 409
Salo, A. 907
Sampurno, Y. 465
Santini, C. 845
Saraya, T. 55
Sarkar, C. K. 39
Sato, S. 71
Sattler, A. 995
Scheffler, A. 751
Schwaha, P. 213
Seacrist, M. 1049
Serebryakov, A. 333
Shaikh, A. 1049
Shang, L. 895, 1075
Shaviv, R. 269
Shek, M. 593
Shellenbarger, Z. 887
Shen, H. 1065
Shen, M. 641
Shen, S. 367
Shen, Z. 15, 27, 117
Shen, Z. 135
Sheu, G. N. 123, 129
Shi, J. 193
Shi, L. 575
Shi, X. 105
Shi, Z. 367
Shiao, D. 491
Shibata, T. 787
Siblerud, P. 689
Sigl, A. 721
Simoen, E. 3, 61
Sinno, T. 959
Soh, Y. 831
Song, S. 281
Song, X. 587
Song, Z. 429, 559
Sopori, B. 1049
Sprey, H. 275
Sriram, M. 269
Srivastava, A. 39
Stimpfl, F. 213
Strandjord, A. J. 751
Sugii, N. 39, 71
Sullivan, S. M. 745
Sullivan, T. D. 205
Sun, F. F. 517
Sun, R. 677
Sun, S. 135
Sun, W. 1043
Sun, W. 641, 645, 659
Sung, J. C. 523, 535
Sung, M. 523
Svensson, B. G. 1013

Sverdlov, V. 213
Swinnen, B. 695
Sy, H. G. 793

Takano, N. 683
Tan, B. 435, 547
Tan, C. 441
Tanaka, H. 727
Tanaka, T. 33
Tanaka, T. 683
Tang, H. 1081
Tang, J. 287
Tang, J. 511
Tang, J. 27
Tang, P. 99
Tanjyo, M. 1059
Tarnowka, A. 453
Tee, P. 669
Teo, H. L. 669
Teramoto, A. 465
Teutsch, T. 751
Tey, T. 187
Tezcan, D. S. 695
Theng, S. 465
Thieme, T. 777
Tolani, V. 299
Tomiyoshi, K. 733
Torfs, T. 707
Trefonas, P. 397
Trojman, L. 651
Tsai, J. 517
Tsaur, S. 129
Tsutsui, K. 39, 71
Tyaginov, S. E. 213

Ukraintsev, V. 169
Umisedo, S. 1059
Une, H. 1059

van Adrichem, P. 357
van Borkulo, J. 837
Van den Bogaert, N. 935
Van Hoof, C. 707
Van Olmen, J. 695
Vanfleteren, J. 707
Varela Pedreira, O. 199
Vartanian, V. 151
Vedantham, R. 529
Verdonck, P. 275
Vereecken, P. M. 845
Verzosa, J. 757
Vijila, C. 869
von Ammon, W. 995
Voronkov, V. V. 945

Wang, B. 15, 27
Wang, C. 511
Wang, D. 397
Wang, F. 381
Wang, H. 895, 1075
Wang, H. 605
Wang, J. 1081
Wang, J. 43
Wang, L. 321, 341
Wang, L. 567
Wang, Q. 49, 93, 141, 901
Wang, R. 99
Wang, S. 441
Wang, S. 491, 511, 635
Wang, X. 83
Wang, X. 373, 381
Wang, X. 641, 645
Wang, X. 875
Wang, Y. 141
Wang, Y. 99
Wang, Y. 875, 1081
Wang, Y. 99
Wang, Y. 43
Wang, Y. 983
Wang, Z. 21
Washburn, C. 419
Webber, S. 575
Wei, F. 321, 341
Wei, H. 587
Wei, S. 117
Weiwei, L. 497
Weldon, M. 567
Wen, R. 479
Wen, T. 575
Wen, X. 231
Wen, Y. 315
Willson, C. 575
Wimplinger, M. 721
Wisse, B. 357
Witvrouw, A. 199
Wong, A. 351
Wu, C. 863
Wu, C. J. 397

Wu, D. 99
Wu, D. 1043
Wu, H. M. 281
Wu, H. 83, 219
Wu, H. 269
Wu, J. G. 281
Wu, L. 935
Wu, P. 321, 341
Wu, T. 435, 547
Wu, W. 269
Wu, Y. 219, 645

Xia, A. 503
Xia, L. 593, 605
Xia, Z. 511
Xiang, D. 243
Xiang, Y. 429
Xiao, D. 83
Xiao, G. 299
Xie, C. Q. 93
Xie, C. 367
Xie, J. 907
Xie, J. 83
Xie, Y. 249
Xie, Z. 429
Xing, C. 601
Xing, C. 677
Xiong, S. 441
Xu, C. 397
Xu, J. 441, 581, 605
Xu, J. 135
Xu, L. 587
Xu, L. 907
Xu, L. 575
Xu, M. 397
Xu, Q. 605
Xu, S. 663
Xu, S. 459
Xu, X. 187
Xu, Z. 287
Xue, B. 67

Yamato, Y. 231
Yan, E. 581
Yan, F. 43
Yan, L. 135
Yang, C. 491, 511, 553
Yang, D. 1001
Yang, F. 43
Yang, H. 611
Yang, J. 219
Yang, S. 123
Yang, S. 763
Yang, Y. 441
Yang, Y. 367
Yang, Y. 881
Yauw, O. 803
Ye, H. 105
Ye, T. C. 93
Ye, W. 593
Yi, C. 43
Yi, S. 763
Yin, C. 33
Yin, H. 875
Yin, X. 645
Yoshikawa, K. 787
You, H. N. 129
Yu, C. 511, 553
Yu, J. 635
Yu, R. R. 851
Yu, S. 291
Yu, X. 67
Yu, Y. 83
Yuan, D. 83
Yuan, J. 135
Yuan, Q. 373, 381

Zakel, E. 751
Zeng, Y. 1001
Zhang, B. 581
Zhang, F. 429
Zhang, H. 641, 645, 659
Zhang, H. 1065
Zhang, J. 677
Zhang, J. 447
Zhang, L. 99
Zhang, P. W. 93
Zhang, P. 575
Zhang, S. 49, 141, 901
Zhang, S. 77, 111
Zhang, S. 529
Zhang, T. 875
Zhang, X. 1001
Zhang, X. 575
Zhang, Z. 559
Zhang, Z. 605
Zhao, D. 581
Zhao, J. 1043

Zhao, L. 659
Zhao, R. 441
Zhao, X. 435
Zhao, Y. 249
Zhong, M. 429
Zhong, X. 757
Zhou, A. 27, 117
Zhou, H. 511
Zhou, H. 601, 677
Zhou, J. 1065
Zhou, P. 281
Zhou, W. 135
Zhou, W. 813
Zhu, C. X. 93, 1071
Zhu, F. 869
Zhu, J. 321, 341
Zhu, Y. 447
Zhu, Z. 321, 341
Zhuang, Y. 465
Zhuge, J. 99
Zoberbier, M. 713
Zubkov, V. 593
Zuo, Q. 141